AN ANTHOLOGY OF PHILOSOPHICAL STUDIES
VOLUME 3

Editor
Patricia Hanna
University of Utah
USA

Editorial Board

Gerald Doppelt
University of California, SanDiego
USA

Yiorgo N. Maniatis
University of Cyprus
Cyprus

Carol Nicholson
Rider University
USA

Donald Poochigian
University of North Dakota
USA

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

Isham Ahmad, International Islamic University, Malaysia

George Bruseker, University of Athens, Greece

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

Effie Papoutsis Kritikos, Northeastern Illinois University, USA

Ian O’Laughlin, University of Idaho, USA

Seung-Kee Lee, Drew University, USA

Joel Martinez, Lewis & Clark College, USA

Mark A. Michael, Austin Peay State University, USA

Chris Onof, Birkbeck College, UK

Elizabeth Schiltz, College of Wooster, USA

Ioannis Trisokkas, University of Warwick, UK
AN ANTHOLOGY OF PHILOSOPHICAL STUDIES
VOLUME 3

Edited by Patricia Hanna
# Table of Contents

1. **Introduction**
   
   *HANNA, PATRICIA*

   **Part 1: History of Philosophy**

   **Ancient and Medieval**

2. **John Duns Scotus on Matter**
   
   *BOLYARD, CHARLES*

3. 'More Familiar to Us’ vs. 'More Familiar Simpliciter'
   
   *HUSSAIN, HAMMAD A.*

4. Thrasymachus and the Relational Conception of Authority
   
   *LONG, RODERICK T.*

5. Matter and Form, Number and Nows: An Analogy for Understanding Aristotle’s Account of Time
   
   *PETERSON, ANNE*

6. Is Machiavelli a Tragic Philosopher?
   
   *PETERSON, URSULA NIKLAS*

   **Modern and 19th Century**

7. The Problems of Hobbes’s Concept of the Sovereign
   
   *HUANG, HSUAN*

8. Towards a Kantian Anthropology: The Question of Human Being in Kant’s *Groundwork* and Second Critique
   
   *McLUCKIE, ALAN*

9. Hegel’s Dialectic as Bridge between Analytic and Continental Philosophy
   
   *PONZER, HOWARD*

10. The Social Nature of Beauty: Kant’s Response to Burke
    
    *VANDENABEELE, BART*

11. Kant on Political and Ethical Communities with Regard to Perpetual Peace
    
    *VAN IMPE, STIJN*

   **20th Century**

12. Foucault: The Cost of Truth and the Price of Transgression
    
    *DIAZ, JULIO CESAR*

13. Craig and His Concept of Eternity: A Critique from the Standpoint of the Kālām
    
    *ERDEM, ENGİN*

14. Remarks on the Poetic Quality of Wittgenstein’s Philosophy
    
    *GITSOULIS, CHRYSOULA*

15. Collingwood and Rorty on the Role of Philosophy
    
    *NICHOLSON, CAROL*
16. An Examination of Richard Rorty's Neo-Pragmatist Educational Views
   NOAPARAST, KHOSROW BAGHERI

17. Reasons and Causes for an Aesthetic Response
   WARD, ANDREW

Part 2: Metaphysics, Epistemology and Philosophy of Language

18. Logically Unknowable Propositions: A Criticism of Tennant’s Three-Partition of Anti-Cartesian Propositions
   CARRARA, MASSIMILIANO & FASSIO, DAVIDE

19. The Chomskyan Revolution: A Category Mistake?
   HANNA, PATRICIA

20. Gödel, Mathematical Knowledge and Skepticism
    McEVOY, MARK

21. Naturalizing Qualia
    PLATCHIAS, DIMITRIS

22. Abstraction and the Interpretation of Quality
    POOCIGIAN, DONALD V.

23. Hegel’s Early Response to Pyrrhonian Skepticism
    TRISOKKAS, IOANNIS

24. Grounding Semantic Contextualism on Epistemological Contextualism
    VASSALLO, NICLA & BIANCHI, CLAUDIA

Part 3: Ethics and Political Philosophy

25. Can Political Liberalism Deliver Equality in the Social-Bases of Self-Respect?
    DOPPELT, GERALD

26. Human Rights, Needs, and Autonomy
    HASSOUN, NICOLE

27. Plato, Moore and Ethical Nonnaturalism
    KYRIACOU, CHRISTOS

28. Socio and Politico – Philosophical Ideas in Theories of Central Asian Thinkers
    LEVINSKAYA, VICTORIA O.

29. The Abolition of War: A Philosophical Imperative
    OBERST, JOACHIM L.

30. Apologies and Moral Harm
    SAVELLOS, ELIAS E.

31. Blackburn’s Supervenience Argument against Moral Realism
    SUZUKI, MAKOTO
List of Contributors

Khosrow Bagheri Noaparast is Professor of Philosophy of Education at the University of Tehran and currently the president of Philosophy of Education Society of Iran (PESI). He is the author of Islamic Education (Tehran, Alhoda, 2001) and a chapter in Advances in Personal Construct Psychology (Neimeyer, R.A. ed., London: JAI Press Inc, 1995). He has also published many articles on philosophy of psychology and philosophy of education.

Claudia Bianchi is Associate Professor at the Philosophy Faculty of the University Vita-Salute San Raffaele, Milan; she teaches Philosophy of Language. She has written many articles and reviews in Italian, French and English: her latest authored work is Pragmatica del linguaggio (Laterza, Roma-Bari 2003, 6th edition 2008), and the most recent edited books are The Semantics/Pragmatics Distinction (Stanford, CSLI, 2004) and Filosofia della comunicazione (Roma-Bari, Laterza, 2005, with Niela Vassallo). She works in the field of Analytic Philosophy. Her main research interest lies in the field of Philosophy of Language and Pragmatics. Other research interests include Epistemology and Philosophy of Science.

Charles Bolyard is an Associate Professor of Philosophy at James Madison University in Harrisonburg, Virginia, USA. He works on medieval epistemology and metaphysics generally, and he has published on the treatments Augustine, Siger of Brabant, Peter Auriol, and John Duns Scotus give of aspects of these issues. His papers have appeared in such journals as Vivarium and the Journal of the History of Philosophy.

Massimiliano Carrara is assistant professor of Logic and Philosophy of Science at the Department of Philosophy of the University of Padova, Italy. He graduated and received his Ph.D. in Philosophy from the University of Padova. His main research interests are in philosophical logic and metaphysics, in particular on the notions of identity and existence.

Julio César Diaz is an Assistant Professor at Chinese Culture University, Taiwan. His most recent work, Being in the Pampas, discusses the reception of Parmenides’s notion of being in Argentine Literature. His article on luck and poesis in ancient Greek thoughts was published in International Studies in Philosophy in 2006. He is now working on Foucault’s epistemology.

Gerald Doppelt is a Professor of Philosophy and Science Studies at University of California, San Diego and also holds the title of UCSD Academic Senate Distinguished Teacher. His research focuses on issues in philosophy of science such as Kuhn, Scientific Rationality, Values in Science, Scientific Progress, and Scientific Realism. His research in Political Philosophy focuses on Rawls, Political Liberalism, Multiculturalism, Feminism, Theories of Social Justice,
the Philosophy of Technology, and Critical Theory of Race. Professor Doppelt teaches large required courses on public issues and the ethics of science, technology, and society.

**Engin Erdem** is Researcher at Divinity School in Ankara University, Turkey. He received his MA and PhD in Philosophy of Religion from the Faculty of Divinity from Ankara University. His philosophical interests lie chiefly in the Nature of Divine Eternity, Philosophy of Time, Medieval Islamic Philosophy, and Kantian ethics especially his later philosophy.

**Davide Fassio** is PhD student at the University of Padova, Italy. He graduated in Philosophy at the University of Turin with a thesis on the Knowability Paradox. The topic of his PhD thesis will be the analysis of logical problems for epistemic notions of truth. His main research interests are in epistemology and epistemic logic.

**Chrysoula Gitsoulis** is an Adjunct Lecturer in Philosophy at City College of the City University of New York. She received a B.A. in Philosophy, B.S. in Mathematics, and Minor in Classics from Rutgers College, New Jersey, in 1991. She earned her doctorate in philosophy at the Graduate Center of the City University of New York in Oct. 2008. Her thesis was on Wittgenstein and the Philosophy of Language, supervised by Paul Horwich, Professor of Philosophy at New York University. Her research interests focus on the History of Philosophy (especially Plato, Aristotle, and Wittgenstein), and Contemporary Moral and Social Issues.

**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, 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 issues in theory of meaning, philosophy of mind, justification, and Chomskyan theoretical linguistics; with Harrison, she is developing a realistic ontology of interpretation which finds it roots in the late writings of Wittgenstein.

**Nicole Hassoun** is an Assistant Professor in Philosophy at Carnegie Mellon University. She is affiliated with Carnegie Mellon's Program on International Relations and the Center for Bioethics and Health Law at the University of Pittsburgh. Nicole writes primarily in political philosophy and ethics and focuses, in particular, on global economic and environmental justice. She is also interested in methodological issues in philosophy and the other social sciences.

**Hsuan Huang** is an Assistant Professor in the Department of English Language and Literature, Chinese Culture University, Taiwan. Her research
interests include the political philosophy around the time of the Enlightenment, and the idea of happiness and the possibility of its realization in civil society. Her most recent publication, “No Safe Place: A Reflection on Humanity” (in *Humanity at the Turning Point: Rethinking Nature, Culture and Freedom*), discusses Kant’s political ideas. In 2008, she presented a paper, “Kant’s Concept of Radical Evil,” at the 9th Global Conference: Perspectives on Evil and Human Wickedness.

**Hammad A. Hussain** is a PhD candidate in philosophy at the University of Oklahoma in Norman, OK, U.S.A. He specializes in Ancient Greek philosophy and Aristotle in particular. He is currently working on his dissertation on Aristotle’s theory of induction. He has presented his paper “‘More Familiar to Us’ vs. ‘More Familiar Simpliciter’” at the 41st meeting of the North Texas Philosophical Association at the University of North Texas in April 2008; at the 3rd International Conference on Philosophy, sponsored by ATINER, in Athens, Greece in June 2008; and at the “Aristotle, Ethics and Science” Conference in Philadelphia, PA, in October 2008.

**Christos Kyriacou** is a Ph.D student at the Philosophy Department of the University of Edinburgh, where he is currently writing a thesis entitled 'The Epistemic Justification Puzzle' under the supervision of Prof. M.Ridge and Dr M.Chrisman. His research interests focus on the semantics of normative sentences (moral and epistemic) and the relevant ramifications in metaphysics, epistemology, mind and action theory. He is also interested in Plato scholarship and in the ways that can be compared and/or contribute in contemporary debates on normativity.

**Victoria O. Levinskaya** obtained her PhD degree in Social and Political Philosophy at the National University of Uzbekistan, where she taught different courses in Philosophy for more than 10 years. Currently she is working at the Westminster International University in Tashkent. She has participated in many International programs including Visiting Fulbright Scholarship program and Fulbright Visiting Specialist program “Direct access to the Muslim world”. Organized a number of International Summer schools for young faculty members of NIS in Uzbekistan, including the “Islam and Civil Society” summer school. She has also taken part in international conferences, and given a number of guest lectures to ten US and European universities; in addition, she has published on different topics of Social, Political and Environmental Philosophy.

**Roderick T. Long** is Associate Professor of Philosophy at Auburn University, Alabama (U.S.A.); author of *Reason and Value: Aristotle versus Rand* (2000) and *Wittgenstein, Austrian Economics, and the Logic of Action* (forthcoming from Routledge); and President of the Molinari Institute. He works chiefly in ancient philosophy, moral and political philosophy, and philosophy of social science, and maintains a blog at [http://aaeblog.net](http://aaeblog.net).
Mark McEvoy is Assistant Professor of Philosophy at Hofstra University in New York. His primary research interests are in philosophy of mathematics and epistemology. His work has been published in several journals, including *Philosophia Mathematica, Dialectica, Metaphilosophy,* and *Philosophical Forum.*

Alan McLuckie is a PhD student in philosophy at Stanford University. He received his B.A. Honours (2005) and M.A. (2007), both in philosophy, from the University of Alberta. His interests are in the history of philosophy more generally, and the intersection between classical philosophy and 18th and 19th century German philosophy in particular.

Carol Nicholson is Professor of Philosophy at Rider University in Lawrenceville, NJ, USA. She has written on political philosophy, philosophy of education, and feminism and is currently working on an article, "Education and the Pragmatic Temperament," for the *Cambridge Companion to Pragmatism.*

Joachim L. Oberst has a joint-appointment in the Religious Studies Program and the Department of Foreign Languages & Literatures at the University of New Mexico. He studied at the University of Marburg (Germany), Goshen College (Indiana, USA), and the University of Heidelberg (Germany) before completing his doctorate in Philosophy at McGill University (Montreal, Canada). His teaching and research interests span 19th & 20th Century Continental Philosophy, Ancient Greek Philosophy, Ethics, and 20th Century Theology. He is presently finishing up a book on Martin Heidegger, *The Bounds of Being: Heidegger on Language and Death—The Existential Connection* under contract with Continuum.

Anne Peterson is a Graduate Presidential Fellow pursuing a Ph.D. in Philosophy at the University of Notre Dame, USA. She received her B.A. from the University of Notre Dame in 2007. Her areas of specialization are Metaphysics, with a particular interest in the problem of universals and modality, and Ancient Philosophy, with a particular interest in Aristotle. She is currently working on the issue of prime matter in Aristotle, both as a problem of Aristotelian scholarship and as it relates to contemporary metaphysics.

Ursula Niklas Peterson is an Associate Professor of Philosophy at Indiana University-Purdue University at Indianapolis, IN, USA. She received her Ph.D. from the University of Warsaw, Poland. She teaches a variety of courses in the history of philosophy (including seminars on Machiavelli and Heidegger), and a course on philosophy and literature. Her research focuses on philosophy and literature, especially tragedy and utopia; hermeneutics; Heidegger: published articles in philosophical journals, such as *History of Philosophy Quarterly* and *Clio;* she is working on a book-length research project on philosophical interpretations of tragedy.
Dimitris Platchias is a Research Fellow in the Department of Philosophy, University of Essex and an honorary research fellow at the Centre for the Study of Perceptual Experience, University of Glasgow. His research centres on issues in the philosophy of mind, psychology and perception and he has published widely in these areas, including recent articles in *Philosophia* and *Teorema*. He is currently editing a book on Representationalism (with Fiona Macpherson, MIT Press).

Howard Ponzer is an Assistant Professor of Philosophy at Molloy College, Rockville Centre, NY, USA. His work has focused primarily on Hegel’s dialectic and its relation to analytic logic, but he has also worked extensively on Plato, Aristotle, and Kant. He has presented papers in the United States and Germany, where he taught for two years in Bremen and Dresden. His most recent publication is entitled, “Reconciliation in Hegel’s Idealism.” He is presently working on an article that explores whether philosophy has a voice in the discussion of gender equality.

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.”

Elias E. Savellos is Professor of Philosophy at the State University of New York, Geneseo, NY, USA. He is the coauthor of *Reasoning and the Law* (Wadsworth, 2001), coeditor of *Supervenience: New Essays* (Cambridge University Press, 1995), and author of various articles on philosophical logic and analytic metaphysics. His current philosophical interests lie in ethics and the philosophy of language.

Makoto Suzuki is a Research Fellow at Nanzan University Institute for Social Ethics, Nagoya, Japan. He received his BA and MA in Letters from Kyoto University, and his PhD in Philosophy from The Ohio State University. His research interests include moral philosophy, moral psychology and history of modern philosophy.

Ioannis Trisokkas is currently an Early Career Fellow in the Warwick Institute of Advanced Studies, Warwick, UK. His research focuses on continental epistemology, philosophy of language and metaphysics. He has published extensively on Hegel and the problem of the expression of truth. He is currently an Early Career Fellow in the Warwick Institute of Advanced Studies.
Bart Vandenabeele is Professor of Aesthetics and Philosophy of Art at Ghent University (Belgium). He is the author and editor of several books in aesthetics, philosophy of art and philosophy of communication. He has published on philosophical aesthetics, the history of philosophy (especially Kant and Schopenhauer) and the philosophy of culture in English, German, Portuguese and Dutch. He has been a Visiting Professor at Missouri Western University and Stellenbosch University (South Africa), and is the editor of A Companion to Schopenhauer (forthcoming with Blackwell’s).

Stijn Van Impe has studied philosophy at Ghent University (Belgium) where he obtained his master degree (summa cum laude) in 2007. He is currently employed as Fellow of the Scientific Research Foundation Flanders at Ghent University and is preparing a PhD on Kant’s realm of ends. He is member of the Centre for German Idealism, the Kant-Gesellschaft Bonn, the Société d’Etudes Kanttiennes de la Langue Française, and the Dutch Association of Aesthetics. He has presented papers on several international conferences both in Europe and the United States and for associations such as the United Kingdom Kant Association and the European Communication Research and Education Association.

Nicla Vassallo is Full Professor of Theoretical Philosophy in the Department of Philosophy at the University of Genova and is part of the teaching staff of the Doctorate in Philosophy. Her primary research and teaching areas are Philosophy of Knowledge and Epistemology. She also has serious research interests in Feminist Philosophy, Metaphysics, Philosophical Naturalism and Scepticism. Her latest sole-authored work is Teoria della conoscenza (Roma-Bari 2003) and her latest co-authored work is Filosofia delle donne (Roma-Bari 2007). Her most recent edited book is Filosofia delle conoscenze (Torino 2006), and her most recent co-edited book is Knowledge, Language, and Interpretation: On the Philosophy of Donald Davidson (Frankfurt, 2008). She regularly writes for the Sunday Cultural Supplement of the newspaper Sole-24 Ore.

Andrew Ward is currently teaching Philosophy at the University of York, UK. He has also taught at the Open University, UK and at the University of Florida at Gainseville, 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 numerous articles in philosophical journals, including Mind, American Philosophical Quarterly, Philosophy and Phenomenological Research, and History of Philosophy Quarterly.
CHAPTER ONE

Introduction

Patricia Hanna

The International Conference on Philosophy sponsored by the Athens Institute for Research and Education (ATINER) has now been held for three years. Each year, we have worked to make it better, and this volume, the third that has been produced in connection with the conferences, shows the results of these efforts.

We now have a five member Editorial Board which works closely with the Editor in putting together the final volume. More importantly, however, we have now established a pool of reviewers which allows us to send each paper to at least two reviewers for their recommendations on the suitability of the papers for publication. It is our aim to treat the papers in a manner which parallels the review standards for professional journals in philosophy; while blind reviewing is not entirely possible, we take this as our model. Starting with Volume II (2007 Conference), we have asked reviewers to base their recommendations on the same standards they would use for reviewing papers for professional journals; for the past two years, approximately 35% of the papers presented at the conference have been accepted for inclusion in these proceedings.

One of the strongest motivations for raising the level of expectation is that this conference is one of a vanishing breed: a small international philosophy conference which is open to all areas of philosophy. It provides an almost unique opportunity for philosophers from all over the world to get together and share ideas with the aim of expanding our understanding of our discipline, and to do so in a venue which allows all of the participants to get to know one another. It is the hope of the Editor and the Editorial Board that the conference will flourish in future years, and that it will draw the best philosophers from every country, regardless of their area or the specific approach or methodology they follow.

The world faces new challenges in terms of shrinking oil supplies, climate change and an uncertain political and economic future. Now, more than ever, bringing philosophers from around the world together to address the most fundamental questions confronting us as a species is needed. At the 2008 conference, held from 2—5 June 2008, 71 philosophers from over 20 countries (including, Belgium, Brazil, Canada, Cyprus, Finland, Germany, Greece, Iran, Italy, Japan, Lebanon, The Netherlands, Portugal, Romania, Singapore, South Africa, Taiwan, Turkey, Uzbekistan, United Kingdom and the USA) met and
talked with one another about the human condition in the shadow of the Parthenon.

We have chosen to organize 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 here 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 breadth and variety of subject matter or the issues brought to the conference itself by the participants, some of whom could not travel to one another's home countries without enormous difficulty.
Part 1

History of Philosophy
Ancient and Medieval
CHAPTER TWO

John Duns Scotus on Matter

Charles Bolyard

In his Questions on the Metaphysics of Aristotle (hereafter, QM),¹ John Duns Scotus presents a roughly Aristotelian view of a roughly Aristotelian problem: he relies upon matter (materia) to explain the process of change. But he also develops his conception of matter in ways that Aristotle never clearly endorses, and perhaps never would. Scotus uses a theological principle inspired by doctrinal events in the year 1277 to interpret and rethink Aristotle’s views. This principle is that because of God’s omnipotence, many things that seem to be necessarily conjoined in nature are not in fact necessarily conjoined.

Scotus employs this general principle to explain, e.g., why the contingent features of things (or accidents, to use Aristotle’s term), such as being short-haired in the case of a dachshund, or being owned by John Coltrane in the case of a saxophone, are separable from things to a much greater degree than Aristotle would allow. In short, Scotus presents an atomistic-leaning, yet still identifiably Aristotelian, account of the world. It is this world that will be my focus; in particular, I will look closely at Scotus’s account of matter, with special attention paid to the reasons he posits its existence (§1), the role matter is meant to play (§2), and its overarching ontological status (§3). I will conclude by discussing some questions raised by his theory (§4), after which I will give a brief closing statement (§5).

Before beginning, it is important to discuss the text I will be exploring. First, note that I will not give a comprehensive account of Scotus’s view of matter. He discusses this topic in his other main works (e.g., in his magnum opus, the Ordinatio), and there has been a great deal of scholarly interest in establishing the ordering of Scotus’s writings. It has recently become clear that the long-held view that the QM predated the Ordinatio cannot be consistently held. The authors of the 1997 critical edition of the QM conclude “that these questions were composed and revised over an extended period of time and that

certain questions stem from a period late in Scotus’s career. The fact that a contrary view was held for so long is due in part to the nature of the work itself: as Stephen Dumont put it, the QM ‘has gained the reputation...as the most notoriously difficult work of a notoriously difficult author’. My aim, then, is to explain the account of matter offered in the QM alone.

There are many ways to think about matter; a few different accounts will be presented here to ease our way into Scotus’s discussion. The first is what we may call quantum matter. On this view, matter comes in discrete chunks, much as the early atomists imagined. These chunks are strong unities, insofar as they cannot lose their individuality without being utterly destroyed. They cannot be further broken down into parts.

The second account we may call massed matter. Here, matter in itself isn’t distinguished into identifiable parts, though it can be divided into parts if one has the appropriate tool. Think of using a knife to divide a large mound of clay into several smaller lumps. In Berkeley’s Three Dialogues discussion of an amalgamated Aristotelian/Lockean theory of matter, he appears to attribute just such a theory to Hylas, the materialist interlocutor of the work.

Finally, we can imagine a combination of the first two accounts. Here too, matter begins as an undifferentiated mass, and it becomes individualized only when it has been separated off from that mass. But unlike the massed matter account, the tool is not removable: the particular thing that separates bits off must remain with the bits in order for the bits to retain their distinctness. The analogy is far from perfect, but imagine the way one separates off a spoonful of soup by lifting it above the bowl from which it was drawn; if the spoon is removed, the spoonful doesn’t retain sufficient unity to remain individualized, and will drop back into the bowl and mix with the rest of the soup in such a way that it’s undistinguishable from that which was already in the bowl. This third account is similar in some ways to the quantum matter account, since the differentiated matter is what one finds in the world. The main difference is that here, the matter isn’t initially differentiated. This differentiation only comes about later. Since I’ll be arguing that Scotus has just such a view, we’ll simply call this Scotistic matter. We’ll return to this basic picture in the third section of the paper.

What Evidence is there for Matter’s Existence?

Scottus holds that our ability to know things in this life is limited. He adheres to the Aristotelian distinction between knowledge of the fact (scientia quia), and knowledge of the reasoned fact (scientia propter quid). Though we can

---

1QM Ed. (Editors’ Introduction), p. xlii.
John Duns Scotus on Matter

speak of knowing that something is the case without knowing everything about it, ‘knowing that’ isn’t knowledge properly speaking; instead, if we truly know something we must know its causes. Early in the QM, Scotus makes this point about metaphysics generally, when he says that ‘metaphysics as knowable to us [in this life]…is necessarily a science of the simple fact’. And this general claim about metaphysics certainly applies to matter. Scotus links fixity and knowability, arguing that matter ‘is mutable, and therefore unknown, i.e., not scientifically knowable’. For him, direct, scientific knowledge of matter in this life is impossible. Instead, the best we can do is to infer its existence indirectly, much in the way that Kant later reasons to the existence of a Ding an sich. So what exactly is the effect that matter produces, and can be known by? ‘It is change that makes matter known’...the generator changes something from one form to another form, and what is changed is called matter. Thus the fact that things change is evidence for the existence of matter. There must be something underlying such changes.

What Role does Matter Play?

Obviously, then, one of matter’s primary roles is to underlie change. At a number of places in the text, Scotus makes the same basic point about this capacity: it is an extremely broad one. As he says, matter has a ‘capacity for infinite real forms’; furthermore, ‘it is in potency to receiving anything whatsoever that can perfect it, and to receiving this in any manner or in any order’. Matter has this potency naturally, but Scotus argues that even seemingly unnatural changes are possible with God’s influence:

...if after having had the form of earth, matter were immediately to be perfected with the form of wine, it would be as naturally perfected by that form [once it is acquired], as it would be, if that wine form had been mediatelty acquired only after another change.

Note that for Scotus, the only odd thing about such a change from earth to wine is the immediacy of it; there is nothing at all odd, in his view, about the same matter existing in wine and (later) in earth. God can skip steps, but in these cases, Scotus still considers the results to be natural ones. Beyond this role in change, however, matter’s influence is severely limited. Whereas thinkers such as Aquinas make matter (at least matter of the right sort) the principle of individuation, Scotus denies any further important

---

1QM V.7, n. 71; Ed. III, p. 507; Trans. I, p. 455.
2QM VII.15., n. 38; Ed. IV, p. 307; Trans. II, p. 264.
3QM VII.5, n. 7; Ed. IV, p. 132; Trans. II, p. 120.
4QM II.4-6, n. 149; Ed. III, p. 279; Trans. I, p. 239.
5QM IX.12, n. 12; Ed. IV, p. 615; Trans. II, p. 553. See also nn. 6, 11, 15, 16, and 17 of the same question.
6QM IX.12, n. 16; Ed. IV, p. 616; Trans. II, p. 554.
functions for it. As he says when comparing matter to the subject of a science in an epistemological context, ‘the subject gives a science its unity, its distinction from other sciences, its order, and its necessary character, none of which matter gives.’ Furthermore, and most famously, Scotus gives multiple arguments against matter as an individuating principle in *QM* VII.13. This is a view to which he has an extremely strong commitment.

There are five key arguments Scotus gives here; only the first one has an overtly theological character. As he says, ‘the first agent [i.e., God] can produce many individuals without matter’: God could produce two (non-material) heavens, for example. This basic claim about immaterial individuals also extends to mathematical objects in his second argument. If matter were the principle of individuation for such objects, then ‘there would not be many distinct individuals simultaneously in the same species, since they do not have distinct matter simultaneously.’ Though he gives no example here, perhaps he would say that if matter were the principle of individuation, then it would be impossible to have two individual triangles, or two instances of the number five. Even so, perhaps Scotus could allow these exceptions, and might still hold that matter could serve as the individuating principle for a limited range of things—namely, for those that are material.

But Scotus denies this possibility as well. His third argument is his simplest, and perhaps his most convincing:

...the same matter numerically that is under the form of one individual can be under the form of another subsequently, and therefore matter is not that which distinguishes two individuals, or that by which this is just this.  

In short, since matter can exist first as part of one individual, and later as part of another (distinct) individual, then matter obviously cannot be the individuating principle. If it were, the individuals would be identical to one another. To return to the example above, the earth that is changed into wine would be identical to the wine it is (substantially) changed into. As Scotus

---

1Dumont argues that some theological background explains Scotus’s rejection of a matter-based principle of individuation. According to Dumont, Scotus is responding to the Condemnation of 1277 and one of its primary authors, Henry of Ghent, since Henry declared that one of the theological problems about angelic individuation resulted from the incorrect Aristotelian view that matter individuates. See Dumont (1995), p. 201.

2 *QM* I.Prologus, n. 31; Ed. III, p. 14; Trans. I, p. 11.

3The order of presentation here is mine, not Scotus’s. Also, note that he gives other arguments as well; the ones that appear here are those that are most relevant to the discussions in the paper.

4 *QM* VII.13, nn. 93-94; Ed. IV, p. 249; Trans. II, pp. 249-250.

5 *QM* VII.13, n. 46; Ed. IV, p. 234; Trans. II, p. 204.

6 *QM* VII.13, n. 40; Ed. IV, p. 233; Trans. II, p. 203.

7Of course, at any give point in time, no two individuals can share the same bit of matter, and in that sense, matter could serve as a distinguishing characteristic. But this is an accidental feature of matter at best, and if this is all that is required, then (e.g.) accidents such as location could serve as individuating principles as well.
makes clear at the end of the passage above, matter provides neither an individual’s individuality (‘that by which this is a this’), nor its distinctness from others of the same species.

He expands on this basic view in his fourth argument, and it is this expansion that provides the main grounds for my overarching interpretation of Scotus’s view in the final section of the paper. Though the explicit context here is an argument against using ‘quantity’ as the principle of individuation, Scotus tells us (n. 40) that this argument works against matter as well. Here’s what he says:

...either quantity has indefinite dimensions or not. But the first is not the case, because the same indeterminate quantity remains in the production of air from fire...Not definite dimensions either...because these follow the form in matter and the individuating quantity is assumed to exist in the matter prior to its reception of the form by making a distinct part of matter which is capable of [receiving] a distinct form...{

As I read the passage, then, Scotus is arguing that matter is either undifferentiated (what I called ‘massed matter’ above), or not. If it is undifferentiated, then every material thing shares the same undifferentiated matter, and thus matter cannot possibly serve as the principle of individuation. If it did, there would only be one individual in existence. On the other hand, if matter is differentiated (that is, if matter is not massed, but is already separated into bits), then whatever it is that does the separation is the true principle of individuation. Using my analogy from the introductory section above, it is the tool that individuates, not the matter lumped together by means of the tool.

So what exactly is this tool? As Scotus tells us in his fifth argument, it is an ‘individual difference’ (differentia individualis), or as it later came to be called, a haecceity.\(^2\) He explains its nature to some degree in n. 124:

This individual difference, because it is not a principle of constituting anything as predicable, but only as subjectable...is not the principle of definition.\(^3\)

In other words, the haecceity breaks matter into bits, preparing these bits for the reception of whatever substantial and accidental forms the matter has.\(^4\) It is

---

\(^1\)QM VII.13, n. 33; Ed. IV, pp. 229-230; Trans. II, p. 200.
\(^2\)Despite the fact that ‘haecceity’ has come to be the standard term used to refer to this aspect of Scotus’s metaphysics, he himself uses the term ‘haecceity’ only on occasion in the QM; he uses other terms as well.
\(^3\)QM VII.13, n. 124; Ed. IV, p. 261; Trans. II, pp. 227-228.
\(^4\)According to Gracia, (who relies primarily on the Ordinatio account of haecceities) the reception of a haecceity in matter is simultaneous with its reception of a substantial or accidental form, but when Scotus discusses the issue, he holds that haecceities are posterior to the things they unite. This is a metaphysical as opposed to a temporal posteriority, however. See Gracia, J. (1996). ‘Individuality and and the Individuating Entity in Scotus’s Ordinatio:
the substantial form that figures into a thing’s definition, not the haecceity. This haecceity is an exceedingly minimalist individuating principle; Scotus tries to be as theoretically economical as he can here.

Now that we have some sense of what matter does (namely, underlies change) and what it doesn’t do (namely, it neither individuates not unifies), we can move towards gaining a fuller understanding of just what it is.

**What is Matter’s Ontological Status?**

As King (2003) and Williams (2007) have recently argued, when one looks at Scotus’s corpus as a whole, one finds him saying that prime matter exists, and furthermore, that it can exist separately from form. Closer inspection of Scotus’s *QM* discussions of matter, however, shows that while it is technically correct to say that he admits the existence of prime matter, what he calls prime matter isn’t as primary as one might think. In addition, at many points in the *QM*, Scotus denies that matter can exist independently. I will explain by presenting a three-fold account of his view.

First, Scotus discusses matter in its purest form (matter qua matter; matter per se; matter taken absolutely). While this kind of matter is infinite, it is utterly undifferentiated, since ‘of itself it is indistinct’. Unsurprisingly, then, Scotus also holds that ‘matter [taken] per se is not individual’, and a fortiori, ‘matter absolutely is not of itself this’—that is, it is not a substance. Finally, Scotus tells us that matter of this sort does not exist independently at all: ‘what is essential to substance is to exist per se separably from others, while not depending on others. But this does not pertain to matter save through form.’

Given this description, many interpreters of Aristotle and Aristotelianism would think that Scotus holds that this first sense of matter is prime matter. But this would be a mistake. For Scotus, prime matter must have some sort of actuality (or as he puts it, ‘positive entity’) in order to serve as the subject of predication. If prime matter is pure potentiality, then ‘it is, simply speaking,
not a being'.¹ that is, it doesn’t exist at all. Though he never gives this argument explicitly, he says enough in other contexts to suggest that if prime matter were so purely potential as to not even register as a being, any change we institute in matter would amount to creation ex nihilo, which for Scotus is reserved for God alone. Given all of this, then, matter in this sense is close to what I called ‘massed matter’ in the introduction; at best, it is subprime matter. As I read Scotus, this subprime matter’s status is similar to that of Scotistic common natures. For Scotus, common natures exist only in things (e.g., Plato’s individual humanity as it exists in Plato himself), or in minds (e.g., the universal concept of humanity when it is thought by a thinker); they never exist independently, and in themselves they are neither universal nor singular. They are ‘indifferent’ to both characterizations. In the same way, subprime matter isn’t singular in itself, and it too—again, when taken in itself—is indifferent to whatever singulars it ultimately becomes a part of.²

The second way Scotus discusses matter is qua prime matter, properly so called (he also terms it ‘neutral matter’, or ‘passive potency’). Matter of this sort consists of subprime matter combined with haecceities. This matter is separated into discrete chunks, but it is minimally differentiated. This allows matter to be something that can serve as the subject of predication, without being determined in any other way.³ Thus prime matter can take on substantial and accidental forms and can lose those forms. Most of the negative characterizations of subprime matter apply here as well—that is, prime matter is not a substance, and does not exist independently. The difference is that chunks of prime matter can be distinct from other such chunks, and thus have some minimal individuality.

The third type of matter is the matter we find in our everyday experience. This is informed matter; it has taken on a substantial or accidental form, and thus has been determined in a certain way. Scotus’s most extended discussion of the way this occurs appears in Book IX. It is best to let him speak for himself here:

...matter’s potency to the being communicated to it by form can be called substantial... [the substantial form] to which this existence pertains per se combines to form one thing with matter which is in potency.⁴

---

¹*QM* VII.5, n. 18; Ed. IV, p. 135; Trans. II, p. 122.
²Gracia discusses a text in the *Ordinatio* in which Scotus explicitly claims that matter is a nature and thus is indifferent to singularity. See Gracia (1996), p. 242. Also, cf. Cross (1995), p. 138, footnote 4. Here, Cross says that ‘Scotus claims unequivocally that prime matter has an essence or quiddity’, and he offers as evidence a reference to *QM* 7.15, nn. 3 and 5-6. But as I read these passages, they offer no such evidence. The reference appears to be garbled in another respect, insofar as the abbreviation ‘Met.’ is supposed to refer to Aristotle’s *Metaphysics*, rather than Scotus’s *QM*.
⁴*QM* IX.1-2, n. 46; Ed. IV, p. 526; Trans. II, p. 470.
And a bit later:

...the potency of matter...is divided properly into natural [e.g., fire’s propensity to rise] and violent [e.g., water’s ability to be heated], and perhaps into neutral or bare potentiality.\footnote{QM IX.12, n. 9; Ed. IV, p. 613; Trans. II, p. 551.}

Scotus explains this further, discussing what he calls ‘obediential potency’. Natural potency, he says, is ‘the first or limited subjection of matter...because nature is determined to something or some things.’\footnote{QM IX.12, n. 13; Ed. IV, p. 615; Trans. II, p. 553.} Obediential potency, on the other hand, is the ‘limitless subjection to an agent... [I]t is in potency to receiving anything whatsoever that can perfect it.’\footnote{QM IX.12, n. 12; Ed. IV, p. 615; Trans. II, p. 553.} He expresses some indecision about how exactly to explain matter’s potency, but it is clear that he wants to make a separation among types of some sort: prime matter has an unlimited capacity for taking on forms, while matter that has been informed is more limited (except for the case of God’s special influence) in the potentialities it can realize. The bit of matter that helps compose my dog, for instance, cannot take on the common nature felinity at the same time that it retains the common nature caninity.

Questions about Scotus’s Theory

The first question is an interpretive one: does Scotus really subscribe to a subprime matter/prime matter division? Evidence that he doesn’t includes the fact that he uses the term ‘matter taken absolutely’ in many ways, and in some contexts, it appears to refer to prime matter, rather than subprime matter. Furthermore, Scotus never makes an explicit division of the sort I attribute to him. In my view, however, Scotus is forced into making such a division, on pain of contradiction. If matter is ‘indistinct’, as he puts it, it cannot at the same time, and in the same way, be distinguished into (predicable) chunks.

Given this basic division, then, another question arises. How thoroughgoing is matter’s potentiality for change? Could a chunk of matter lose its haecceity and drop back into being subprime matter alone? Scotus doesn’t discuss this issue either, but given his frequent expression of the view that matter exists only as part of a compound, he would be unlikely to say that matter can exist without any haecceity at all. This still leaves open the question of whether matter can obtain a new haecceity in an instantaneous switch. He certainly has no problems having instantaneous substantial changes (again, he allows that God could immediately change earth into wine) with respect to an individuated chunk of matter. And when he discusses common natures, he similarly has no problem with such a nature being contracted to one individual (say, Plato) for a time, and then losing that connection when Plato ceases to exist. If a haecceity

\footnote{QM IX.12, n. 9; Ed. IV, p. 613; Trans. II, p. 551.}
\footnote{QM IX.12, n. 13; Ed. IV, p. 615; Trans. II, p. 553.}
\footnote{QM IX.12, n. 12; Ed. IV, p. 615; Trans. II, p. 553.}
is really something superadded to subprime matter, then such haecceity-switches seem to be perfectly consistent with his views.

This leads to a more fundamental question. If, as I argue here, matter-as-differentiated-by-haecceities is really what Scotus means by prime matter, then it appears that haecceities are doing all of the theoretical work. So if this is the case, why introduce the notion of matter at all? Why not simply have haecceities as the ultimate grounding of change, thus dispensing with what he admits to be an unknowable material substrate? In other sections of the *QM*, Scotus admits that forms sometimes exist without matter. He cites accidental forms of sound in particular,\(^1\) and since sounds obviously change over time, getting weaker as they move farther from their sources, and since God is omnipotent, it is clear that some changes occur without any matter that exists as the subject of such changes. Given this, I see no good philosophical reasons for Scotus to reject this move. But because of his theological commitments—namely, his view that bodies are inherently material, and that the return to one’s body at the end of the world is an important part of Christian doctrine—Scotus would be unlikely to get rid of matter altogether.\(^2\)

And finally, a comparative question arises. According to Aquinas, designated matter—that is, matter-plus-quantity—is the principle of individuation. If so, then is Scotus’s theory a significant improvement over Aquinas’s? All Scotus has done is to posit the existence of a haecceity, which he claims is not an accident, to play the role of quantity in Aquinas. If in fact Scotus’s primary reason for rejecting a quantity-plus-matter account of individuation is due to the accidental nature of quantity, and thus to its metaphysical posteriority to and dependence upon the individual matter it is supposed to individuate, then wouldn’t the same argument apply to a haecceity’s supposed posteriority to matter? In my view, Scotus’s account is an improvement. On my reading, since subprime matter cannot exist independently, and since haecceities are not dependent beings in the way that accidents are, we don’t run into the same cart-before-the-horse problems we do with quantity: whereas quantity, qua accident, is dependent upon the independent, already-differentiated existence of the subject it is supposed to individuate, haecceities are not dependent in this way. Though haecceities never exist outside of combination with matter, they are not dependent on matter.\(^3\)

\(^1\) *QM* VII.4, nn. 17-18; Ed. IV, p. 125; Trans. II, p. 114.
\(^3\) For a clear, comparative discussion of the general accounts of matter offered by Aquinas, Scotus, and Ockham, see Adams (1987), vol. II, pp. 633-695. As Adams explains their positions, Aquinas has a variation of what I call a ‘massed matter’ account, and Ockham has a position that is somewhere between that of Scotus and what I call a ‘quantum matter’ account.
Conclusion

Scotus’s QM account of matter seeks to avoid some of the problems faced by atomistic quantum matter views: it allows for greater unity of the material individual, since the material atoms of which they are composed cannot exist on their own, and it also provides a better explanation of how the individuals of a given species can share a common nature while simultaneously retaining their own individuality. It also avoids the problems of an utterly undifferentiated, massed matter view, since such a view makes individuation problematic, and since something that is utterly undifferentiated and uncharacterized comes uncomfortably close to thoroughgoing non-existence. As Paul Vincent Spade succinctly put it, there is no metaphysical free lunch: if something does theoretical work, it must be granted some (positive) metaphysical status.

Scotus’s view is not without problems, primarily with respect to a haecceity’s metaphysical posteriority to common natures (and presumably, to matter as well). Unfortunately, Scotus is much clearer about the relation between a common nature and a haecceity than he is about the relation between matter and a haecceity. And being less clear than that perplexing explanation is an achievement in itself.
‘More Familiar to Us’ vs. ‘More Familiar Simpliciter’

Hammad A. Hussain

In Posterior Analytics A.2 at 71b35-72a6, Aristotle distinguished between what is more familiar ‘to us’ and more familiar simpliciter (haplos). According to the common reading of this passage, Aristotle considers what is most universal (katholou) to be most familiar haplos and least familiar to us and considers the particulars (ta hekasta) to be most familiar to us, and least familiar haplos. This, however, appears to conflict with A.13 at 78a22-b11. In regard to the pair of deductions on the planets, Aristotle implies that ‘non-twinckling’ is more familiar (apparently, more familiar to us) than ‘near.’ Similarly, in regard to the deduction on the moon, he implies that ‘waxes in this way’ is more familiar (again, apparently to us) than ‘spherical.’ However, in regard to the deductions on the planets (which seem to assume a context of heavenly bodies), ‘non-twinckling’ does not seem to be less universal and more particular than ‘near’; the two terms seem to be of equal breadth. Similarly, in regard to the deductions on the moon, ‘waxes in this way’ does not seem to be less universal and more particular than ‘spherical’; the two terms, again, seem to be of equal breadth.

To resolve this apparent conflict between A.2 and A.13, I propose, as a hypothesis, that the correct interpretation of ‘katholou’ as it appears in the passage in A.2 is different from that of the common reading. More positively, I propose the hypothesis that ‘katholou’ as it appears in A.2 means ‘predicable of the subject as a whole,’ where ‘as a whole’ does not mean ‘universally.’ Roughly speaking, I hold that ‘as a whole’ should be taken to mean ‘essentially.’ As I will argue later in greater detail, there are two pieces of evidence that support this hypothesis. The first is that it preserves the general meaning of ‘katholou’ (which is ‘in general’ or ‘on the whole’). The second, as I will argue, is that it resolves the tension between A.2 and A.13. (I maintain that, other things being equal, it is better, on the ground of charity, if a reading of a passage can render A.2 and A.13 mutually consistent than if the reading cannot.) The value of my hypothesis, if it can indeed resolve the tension between A.2 and A.13 and be shown to be correct, is that it will bring us closer...
to a reading of the *Posterior Analytics* in general that can be shown to be correct on the basis of textual evidence.

In connection with this, we should note two passages, to which I will refer repeatedly, from *Posterior Analytics*. The first is from A.2 and makes the distinction between ‘more familiar to us’ and ‘more familiar *simpliciter.*’ I will refer to it as ‘Passage 1.’ Here is Passage 1, with ‘*katholou*’ and ‘*ta hekasta*’ translated in the standard way:

\[
\text{Things are prior and more familiar in two ways; for it is not the same to be prior by nature and prior in relation to us, nor to be more familiar and more familiar to us. I call prior and more familiar in relation to us items which are nearer to perception, and prior and more familiar simpliciter items which are further away. What is most universal is farthest away and the particulars are nearest—these are opposite to each other.}^1
\]

The second passage is from A.13. It states how Aristotle thinks a ‘demonstration of the fact’ differs from a ‘demonstration of the reason why’ (which, outside of A.13, he seems to refer to merely as a ‘demonstration’). I will refer to it as Passage 2: ‘...it proceeds not through the explanation but through the more familiar of the converting terms. For there is no reason why the non-explanatory counter-predicated term should not sometimes be more familiar, so that the demonstration will proceed through this term.’^2

To illustrate Passage 2, consider the pair of deductions on the moon in A.13. The ‘demonstration of the reason why’ is: all of what is spherical waxes in this way; the moon is spherical; so the moon waxes in this way. The corresponding ‘demonstration of the fact’ is: all of what waxes in this way is spherical; the moon waxes in this way; so, the moon is spherical. The ‘non-explanatory counter-predicated term’ is, it seems, ‘waxes in this way.’ ‘Waxes in this way,’ it seems, is more familiar (evidently, to us), than ‘spherical,’ and the demonstration of the fact proceeds through it as the middle term.

Before I turn to supporting my hypothesis, we should note two other things in regard to how I think Chapters A.2 and A.13 should be read.

First, I believe that when Aristotle mentions something being more familiar (to us or *haplos*), he is speaking neither simply of a propositional item nor simply of a term, but rather of a term in relation to (i.e. predicated of) another term. The evidence for this belief is Passage 2 from A.13 in connection with the examples in A.13 of demonstrations of the fact. If Aristotle believed that ‘non-twinkling’ is more familiar to us (easier for us to perceive) than ‘near,’

---

2Aristotle (2002), 20, 78a35-9. There is a question about whether the two ‘demonstrations of fact’ are intended to be examples of this way of failing to be ‘demonstrations of the reason why.’ Whatever maybe the case, I treat them as cases where the ‘non-explanatory counter-predicated terms’ (emphasis added) are more familiar (to us), because they do, in fact, seem to be cases of that.
More Familiar to Us’ vs. ‘More Familiar Simpliciter’

This can only make sense if these are in relation to ‘planets.’ If ‘non-twinkling’ and ‘near’ are not understood in relation to ‘planets’ but only as single terms, then they would seem to be equally familiar to us, because we can easily perceive many things that are non-twinkling, as well as many things that are near. But this is not the case if these two terms are understood in relation to ‘planets.’ We can perceive that all the planets are non-twinkling, but we have to infer that all the planets are near (assuming we have no special instruments). A similar observation can be made about the example of the demonstration of the fact about the moon. If Aristotle believed that ‘waxes in this way’ is more familiar to us (‘nearer to perception’) than ‘spherical,’ this can only make sense if these are in relation to ‘the moon.’ If ‘waxes in this way’ and ‘spherical’ are not understood in relation to ‘the moon’ but only as single terms, then ‘spherical’ would seem to be more familiar to us, because we can easily perceive many spherical objects, but not many things that ‘wax in this way.’ But this is not the case if these two terms are understood in relation to ‘the moon.’ We can perceive (i.e. it is more familiar to us) that the moon waxes in this way, but we have to infer that the moon is spherical, on the basis of the fact (the premise) that it ‘waxes in this way’ (again, assuming we have no special instruments).

Second, when Aristotle uses the phrase ‘nearer to perception’ (to state in Passage 1 what he means by ‘more familiar to us’), I take him to mean simply ‘easier to perceive.’ Similarly, by ‘farther from perception’ (and, as such, ‘less familiar to us’), I take him to mean simply ‘more difficult, in some sense, to perceive,’ whether that difficulty is the result of the object being far away, microscopic, or of some other cause. Such a difficulty may require the object to be grasped by one or more inference rather than by direct perception. (I intend my phrase ‘more difficult to perceive’ here to subsume what is more difficult to perceive in the sense of being impossible to directly perceive.) This view in regard to what Aristotle means by ‘nearer to perception’ and ‘farther from perception’ can be regarded as a ‘sub-hypothesis’ used in connection with my main hypothesis. I will argue that my sub-hypothesis should be adopted because it can, in connection with my main hypothesis, resolve the tension between A.2 and A.13.

I will structure my discussion as follows. In Part II, I will argue that my hypothesis is supported by the fact that it preserves the general meaning of ‘katholou,’ and by the fact that it resolves the tension between A.2 and A.13. In Part III, I will consider and answer several objections to my proposal. Finally, in Part IV, I will state my conclusion.

Evidence Supporting the Hypothesis

My hypothesis, again, is simply that katholou in A.2 means ‘predicated of the subject as a whole.’ The first piece of evidence to support this hypothesis is simply that it preserves the general meaning of ‘katholou,’ which is ‘on the
whole’ or ‘in general.’ If ‘katholou’ has a general meaning of ‘on the whole,’ then it is conceivable for Aristotle to use it in certain contexts to mean ‘predicable of the subject as a whole.’

The second piece of evidence to support the hypothesis is that it resolves the tension between A.2 and A.13. How does the hypothesis do this? Observe that the tension, as I have presented it in Part I above, is specifically due to the fact that, given the common interpretation of Passage 1, it claims that a thing’s being less familiar to us (and more familiar haplos) is proportional to its universality (i.e. to its breadth or generality). But given my hypothesis, a term in relation to another is less familiar to us (and more familiar haplos) the more it is predicated of the subject ‘as a whole’ (i.e. essentially). A term’s being predicated of the subject more essentially than another is not a matter of the term being more broad or general.

In connection with Passage 2, consider again the pair of deductions in A.13, on the planets:

I.A. ‘Demonstration of the Fact’

1. All of what is non-twinkling is near.
2. All the planets are non-twinkling.
3. So, all the planets are near.

I.B. ‘Demonstration of the Reason Why’

1. All of what is near is non-twinkling.
2. All the planets are near.
3. So, all the planets are non-twinkling.

The two converting terms are ‘near’ and ‘non-twinkling.’ As I have mentioned in Part I, if we have no special instruments (as Aristotle did not), ‘non-twinkling’ (in relation to the planets) is ‘nearer to perception’ (and as such, more familiar to us) than ‘near’ (in relation to the planets). This claim is consistent with Passage 2.

Now, assume my hypothesis and sub-hypothesis. The claim, implied by Passage 2, that Deduction I.A. proceeds through the ‘more familiar’ (apparently, to us) ‘counter-predicated term,’ namely ‘non-twinkling,’ is consistent with Passage 1. For it is easy to take Aristotle to consider ‘near’ to be predicated of ‘planets’ more essentially, i.e. in a more katholou way, than ‘non-twinkling.’ For we can, it seems, take Aristotle to consider the nearness of the planets (in comparison to the stars) as part of their essence (or ‘what it is’), whereas their non-twinkling would only be a consequence of that essence (as indicated by Deduction I.B.).

Now, in connection with Passage 2, consider again the pair of deductions from A.13, on the moon:

---

1Liddell and Scott, eds. (1940).
II.A. ‘Demonstration of the Fact’

1. All of what waxes in this way is spherical.
2. The moon waxes in this way.
3. So, the moon is spherical.

II.B. ‘Demonstration of the Reason Why’

1. All of what is spherical waxes in this way.
2. The moon is spherical.
3. So, the moon waxes in this way.

The two converting terms are ‘spherical’ and ‘waxes in this way.’ As I have mentioned in Part I, if we have no special instruments (as Aristotle did not), ‘waxes in this way’ (in relation to the moon) is ‘nearer to perception’ (and as such, more familiar to us) than ‘spherical’ (in relation to the moon). This claim is consistent with Passage 2.

Now, assume my hypothesis and sub-hypothesis. The claim, implied by Passage 2, that Deduction II.A. proceeds through the ‘more familiar counter-predicated term’ (more familiar in relation to us), namely ‘waxes in this way,’ is perfectly consistent with Passage 1. For it is easy to take Aristotle to consider ‘spherical’ to be predicated of ‘moon’ more essentially, i.e. in a more katholou way, than ‘waxes in this way.’ For we can, it seems, take Aristotle to consider the spherical shape of the moon as part of its essence (or ‘what it is’), whereas its waxing ‘in this way’ would only be a consequence of that essence (as indicated by Deduction II.B. as a demonstration of the reason why).

Objections and Responses

For what follows, I will consider and respond to four objections.

First, one might object that, as with the common interpretation of A.2, ‘katholou,’ changes its meaning in A.4 given my interpretation of A.2. Given this objection, my interpretation of A.2 shares a weakness with the common interpretation.

My response is that given my hypothesis, the meaning of ‘katholou’ A.2 can be viewed as being consistent with the meaning of ‘katholou’ in A.4. If we take ‘katholou’ in A.2 to mean ‘predicable of the subject as a whole,’ where ‘as a whole’ means ‘essentially,’ we can take A.4 to further specify what it means for a term to be predicable of the subject essentially. Taken as such, what does A.4 claim a katholou predication—a predication of the subject essentially—specifically consists in? According to A.4, a kata pantos predication is one that holds of all past, present and future instances of a subject term. A kath’auto predication in one sense (which I will refer to as KA1) is one in which the predicated term is part of the account of the ‘what it is’ of the subject.¹ A

¹Aristotle (2002), 7, 73a35-8.
kath’auto predication in another sense (which I will refer to as KA2) is one in which the subject term is part of the account of the ‘what it is’ of the predicated term. The resulting picture, in A.4, of the conditions of a katholou predication is that it is a predication of the subject kata pantos, kath’auto in either of those senses, and as such.

Let us now consider how such a further specification from A.4 of what a katholou predication is would apply to the pairs of deductions (Pairs I and II) from A.13.

Let us consider Deductions I.A. and B. on the planets. It seems that being near is likely in the account (according to Aristotle) of what it is to be a planet (as it seems this is at least part of what would distinguish planets from stars). ‘Near,’ in other words, would be predicated of the planets KA1. And ‘near’ would hold kata pantos of the planets. ‘Near’ (predicated of the planets) is katholou, whereas ‘non-twinkling’ is not. For ‘non-twinkling’ (we could easily think) would not, for Aristotle, be part of the ‘what it is’ of the planets, nor would ‘planets’ be part of the ‘what it is’ of non-twinkling. The planets’ being near, for Aristotle, is ‘farther from perception’ (and is more familiar haplos) than being ‘non-twinkling’ (which is more familiar to us). It is (according to Aristotle) explanatory of the planets being non-twinkling. The demonstration of the planets being non-twinkling (Deduction I.B.), accordingly, proceeds from ‘near’ as the middle term.

Let us consider Deductions II. A. and B. on the moon. ‘Spherical’ for Aristotle would be predicated kata pantos of the moon. Further, it apparently would be predicated KA1 of the moon, since it seems that it would be part of the ‘what it is’ of the moon. ‘Spherical’ accordingly would hold katholou of the moon. ‘Waxes in this way’ (in regard to the moon) would not, it seems, be katholou. For it seems that ‘waxes in this way,’ would not, for Aristotle, be part of the ‘what it is’ of the moon, nor would ‘moon’ be part of the ‘what it is’ of ‘waxes in this way.’ ‘Waxes in this way’ (in regard to the moon) is more familiar to us, whereas ‘spherical’ (in regard to the moon) is more familiar haplos. It is, for Aristotle, explanatory of moon ‘waxing in this way.’ Accordingly, the demonstration of the moon waxing ‘in this way’ proceeds through ‘spherical’ as the middle term.

One may note that A.4 seems to require that both premises of a demonstration (of the reason why, in the terms of A.13) involve katholou predications. So each of the two demonstrations of the reason why in A.13, i.e. Deductions I.B. and II.B., must have katholou predications in both of its premises. I have suggested that Premises I.B.2 and II.B.2 involve katholou predications. Can the same be said for Premise I.B.1 (i.e., all of what is near is non-twinkling) and Premise II.B.1 (i.e., all of what is spherical waxes in this

---

1Aristotle (2002), 7, 73a38-b2.
2Aristotle (2002), 6-8. Aristotle mentions this last requirement for a predication to be katholou: that it hold of the subject ‘as such.’ See ibid., 8 (73b25-9). For my purposes in this paper, I follow Jonathan Barnes’ suggestion in ibid., 118-9 (which is one of two suggestions he gives) of treating an ‘as such’ predication as logically equivalent to a kath’auto predication.
way)? I suggest that the same can be said. In regard to Premise I.B.1, the premise itself seems to predicate ‘non-twinkling’ of near things *kata pantos*. Could ‘non-twinkling’ for Aristotle be predicated of near things *kath’auta* in some way? It seems that certain things at a distance from us, but nearer than stars, could be seen by us as twinkling. An example could be a torch held by a person at night, at a distance. As the person walks closer, the torch no longer twinkles. Aristotle must have been aware of this, and could very well have believed that the correct definition (i.e. account of the ‘what it is’) of ‘twinkling’ is in terms of distance (i.e. near-ness and farness). In the deductions about planets not twinkling, it seems that he is assuming a context of heavenly bodies, for otherwise the demonstration of the planets not twinkling would not make sense; a torch held by a person at night can be nearer than any of the heavenly planets, and yet twinkle. On this basis, we can consider Aristotle to have held that ‘non-twinkling’ (assuming a context of heavenly bodies) is predicated of near things KA2. Accordingly, ‘non-twinkling’ would be predicated of near things *katholou*.

In regard to Premise II.B.1, by ‘waxes in this way,’ it seems that Aristotle is ‘pointing’ to specifically the way spherical things (given certain circumstances) seem to wax. Accordingly, ‘waxes in this way’ would be predicated of spherical things *kata pantos*, and KA2. Thus, ‘waxes in this way’ would be predicated of spherical things *katholou*.

Note that the fact that the premises in each of the two demonstrations of the reason why involve *katholou* predications does not mean that the conclusion must involve a *katholou* predication. Or at least the conclusion need not involve a predication that is *katholou* in an immediate sense (and can involve a predication that is *katholou*, but less so than the predications of the premises—see my answer to the second objection below).¹

We can thus see how, given my interpretation of A.2., the meaning of ‘*katholou*’ in A.2 is consistent with the meaning of ‘*katholou*’ in A.4, A.2 is consistent with A.13, and A.13 can be made consistent with A.4.

A second objection that might be raised is in connection with my response to the first. If A.4 is taken to further specify what a ‘*katholou*’ predication consists in, it does not seem as though we can speak of ‘degrees’ of being *katholou* in regard to a predication. A predication either is or is not *katholou*, but is never *katholou* ‘to a degree.’ But A.2 clearly implies that there are such degrees. For, as Passage 1 indicates, A.2 speaks of what is ‘most *katholou*.’ [Emphasis is added.²]

¹Further, note that even if we grant that in the two pairs of deductions each premise and conclusion, to some degree, involves a *katholou* predication, it would not follow that all of those deductions become demonstrations of the reason why. Deductions I.B. and II.B. are demonstrations of the reason why according to Aristotle because the middle in each of those two are explainatory. (See Aristotle (2002), 20, 78a36-b3, 78b6-11.) By contrast, deductions I.A. and II.A., even if we take them to involve all *katholou* predications, are not demonstrations of the reason why because they do not proceed through explainatory middles. (See ibid.)

²Aristotle (1960), 32-3, 72a4-6.
I suggest that in A.2 Aristotle means ‘most katholou’ in sense of ‘most’ we often use when we speak of a claim being ‘most certain.’ Such a claim is the closest to being, or uniquely is, absolutely certain, in comparison to other claims. A predication is ‘most’ katholou if it is closest to being, or (in the context of what it is being compared to) uniquely is, simply katholou. A predication is more katholou than another if it is closer to being simply katholou. What does it mean to be closer to being simply katholou? It means, assuming that it is kata pantos, that it is closer to being KA1 or KA2. What would it mean for a predication to be closer to being KA1 or KA2 than another predication? It means that the predication would be KA1 or KA2, or closer in terms of logical inferences to the relevant ‘what it is’ account.

Given this account of what it means for a predication to be ‘more katholou’ than another, we can legitimately speak of degrees of being katholou in regard to a predication, and things being ‘most’ katholou, as A.2 does.

In short, given my hypothesis, the fact that A.2 speaks of what is ‘most katholou’ does not undermine the consistency between A.2 and A.4, nor between A.2 and A.13.

A third objection that might be raised is that ‘katholou’ in A.2 is contrasted with ‘ta hekasta,’ which means ‘particulars.’ So, ‘katholou’ in A.2 must mean the opposite of ‘particular,’ namely ‘universal,’ and not what I suggest it means.

I suggest that ‘ta hekasta’ in A.2 does not have to be taken to mean ‘particulars.’ The most general meaning of ‘ho hekastos,’ according to the Lidell-Scott entry, is ‘each’ or ‘every one.’ Given this general meaning, it is conceivable that Aristotle, in A.2 used ‘ta hekasta’ to mean ‘perceivable particular attributes of the subject.’ In contemporary terms, this would be ‘perceivable tropes of the subject.’

In regard to Deductions I.A. and B. from A.13, I argued that ‘non-twinkling’ (in relation to the planets) is ‘nearer to perception’—more familiar to us—than ‘near’ (in relation to the planets). I suggested that Aristotle takes ‘near’ to be predicated in an (immediately) katholou way of the planets. I now suggest that he takes ‘non-twinkling’ in relation to the planets to refer to ta hekasta, i.e. to perceivable particular attributes of the planets. These attributes are each planet’s (assuming only the ones that were known to Aristotle) particular attribute of not twinkling, each of which attribute is perceivable on a clear night (without special instruments). Ta hekasta, the particular attributes of non-twinking of individual planets, are ‘nearer to perception’—more familiar to us—than what is more katholou here, i.e. ‘near’ as predicated of the planets.

In regard to Deductions II.A. and B. from A.13, I argued that ‘waxes in this way’ (in relation to the moon) is ‘nearer to perception’—more familiar to us—than ‘spherical’ (in relation to the moon). I suggested that Aristotle takes ‘spherical’ to be predicated in an (immediately) katholou way of the moon. I now suggest that he takes ‘waxes in this way’ predicated of the moon to be inductively inferred from ta hekasta, i.e. from perceivable particular attributes.

---

1Liddell and Scott, eds. (1940).
of the moon. Each of these attributes is a lighted part of the moon’s particular shape on a given night, each is perceivable on a clear night (without special instruments), and from these we can induce that the moon waxes ‘in this way.’ Ta hekasta, the various visible shapes on different nights, of the lighted part of the moon, are ‘nearer to perception’—more familiar to us—than ‘waxes in this way’ in relation to the moon, which, in turn, is more familiar to us than what is more katholou here, i.e. ‘spherical’ in relation to the moon.¹

A fourth objection that might be raised would be in connection with the response to the third above. With the common reading of Passage 1, ‘katholou’ is taken to mean ‘universal’ and ‘ta hekasta’ ‘particulars.’ With this reading, there is a clear contrast between ‘katholou’ and ‘ta hekasta.’ Such a contrast fits well with the fact that Passage 1 regards what is katholou and ta hekasta as opposites. If, however, we interpret ‘katholou’ to mean ‘predicable of the subject as a whole’ and ‘ta hekasta’ to mean ‘perceivable particular attributes of the subject,’ as I suggest, there is no such clear contrast.

To answer this objection, I will offer another suggestion. My suggestion is that Aristotle believed that there is a contrast. Recall that, in regard to ‘katholou’ meaning ‘predicable of the subject as a whole,’ I said that ‘as a whole’ should be taken to mean ‘essentially’ (rather than ‘universally’). What is more essential, for Aristotle, I suggest, is more difficult to perceive (farther from perception), and has to be inferred. What is ‘nearer to perception,’ on the other hand, is not what is more essential, but may be the basis on which we infer and grasp what is more essential. This, I suggest, is what Aristotle had in mind, when writing the last sentence of Passage 1.

Conclusion

Aristotle seemed to believe that what is more essential (katholou) in regard to a subject is more difficult to perceive, and that perceivable tropes (ta hekasta) of a subject are easier to perceive and less essential, but data on the basis of which we can arrive at what is more essential. It seems that this is the contrast that Aristotle had in mind between what is more katholou and ta hekasta. The evidence is that such a contrast is consistent with the general meanings of ‘katholou’ and ‘ta hekasta,’ and resolves the tension between A.2 and A.13.

Bibliography


¹In A.4, Aristotle says that what is not predicated katholou is incidental (sumbebekos). This does not contradict Aristotle’s taking what is not predicated of a subject katholou (to any degree) to be of ta hekasta, i.e. perceivable particular attributes of the subject.

Thrasymachus and the Relational Conception of Authority

Roderick T. Long

Thrasymachus, in the first book of Plato’s *Republic*, defines justice as the interest of the stronger:

*The just is nothing other than the advantage of the stronger.* (338c)
*The just is the same thing everywhere – the advantage of the stronger.* (339a)

And since in each city the rulers are the stronger, Thrasymachus apparently treats this definition as equivalent to defining justice as the interest of the rulers:

*Is this not what exercises power in each city – the ruling party?* (338d)
*This, then, sir, is what I say the just is, alike in all cities – the advantage of the established rule.* (338e-339a)

If we follow Thrasymachus in identifying the stronger with the rulers, then Thrasymachus’s definition appears to entail P1:

- **P1.** S acts justly iff S acts in the interest of the stronger/rulers.

And P1 in turn evidently entails both P2 and P3:

- **P2.** The weaker/ruled act justly iff the weaker/ruled act in the interest of the stronger/rulers.
- **P3.** The stronger/rulers act justly iff the stronger/rulers act in the interest of the stronger/rulers.

Does Thrasymachus indeed endorse P2 and P3? P2 he certainly accepts, agreeing that ‘to obey the rulers is just’ (339b) and that ‘those who are ruled do what is to the advantage of the one who is stronger.’ (343c)
But P3 he equally clearly rejects.¹ Nowhere does Thrasymachus describe the rulers who rule in their own interest as just. On the contrary, he repeatedly describes those who rule in their own interest as unjust, explaining (344b-c) that while a common criminal ‘commits injustice by the part’ (kata merē), a political ruler ‘has committed injustice by the whole’ (tēn holēn). And consider these further passages:

Injustice ... rules over those who are truly simpletons and just. (343c)
When each of them holds some ruling office, a just person ... gains no advantage .... Of the unjust man the opposite is true in every respect [i.e., the unjust man does gain advantage by ruling, contrary to P3]. (343e)
The most superlative form of injustice, the one that makes the perpetrator of injustice happiest ... is tyranny. (344a)
Injustice, when it is sufficient in extent [meaning, as the context makes clear, when it achieves political power], is mightier, freer, and more masterful than justice. (344c)
Those who commit injustice in a complete manner ... are able to subject cities and nations of people to themselves. (348d)
Would be unjust for a city to undertake to enslave other cities unjustly? ... That is what the best city will do most of all, the one that is most completely unjust. (351a-b)

In short, while one might expect Thrasymachus’s definition of justice to imply that when the stronger/rulers act in their own interest, they are being just, Thrasymachus apparently says just the opposite – that when the stronger/rulers act in their own interest, they are being unjust. Thrasymachus thus rejects P3 in favour of its opposite, P4:

- **P4. The stronger/rulers act unjustly iff the stronger/rulers act in the interest of the stronger/rulers.**

Yet how is P4 to be reconciled with P1, Thrasymachus’s purported definition?

Should we even expect consistency here? There’s a common tendency for interpreters to see Thrasymachus as muddled; Annas (1981), for example, declares that ‘his statements certainly do not add up to a consistent position’

¹Some interpreters present Thrasymachus’s position as though he accepted P3. Sayers (1999), for example, takes Thrasymachus to view justice as ‘the ideology of the ruling class,’ so that ‘[p]eople from other classes have no reason to act justly’ (13; emphasis added) – as though the rulers themselves do have such reason; while Rosen (2005) describes Thrasymachus as holding that ‘the strongest have the right to take away the benefits of the less strong’ (39; emphasis added) – which strongly suggests, contrary to Thrasymachus’s explicit statement, that the strongest are acting justly in doing so.
Thrasymachus and the Relational Conception of Authority

(37), while White (1979) suggests that the ‘confused manner’ in which Thrasymachus presents his position indicates ‘Plato’s opinion that those who hold such positions do not generally articulate them very well.’ (65) Barney (2006) thinks Thrasymachus never settles whether the rulers are supposed to be just or unjust when they rule in their own self-interest because he is ‘not giving a definition of justice but rather debunking it by pointing out the standard effects ....’ (45) For Cross and Woozley (1979), Thrasymachus ‘is not being perfectly consistent,’ but ‘has advanced two different criteria of justice,’ the interest of the stronger and the interest of the other, ‘without appreciating that they do not necessarily coincide. In the case of the weaker they do coincide, but in the case of the stronger they do not.’ (41) Dorter (2006) likewise notes that because ‘if we do good for someone weaker than ourselves it will be just according to the second criterion, the good of another, but not the first, the advantage of the stronger, these ‘cannot both be definitions of justice’ (34) – though Dorter attempts to rescue Thrasymachus from inconsistency by taking only the first claim to be Thrasymachus’s true definition. Weiss (2007) suggests that Thrasymachus must have been ‘unaware’ (97) of shifting from justice as obedience to the rulers to justice as deference to others’ interests generally, while Everson (1998) complains that since the enactments of the tyrant are the only standard of justice that Thrasymachus has offered us, his insistence that ‘the tyrant is unjust because he ignores the interests of other people’ is ‘nonsensical.’ (115-16) In fact most interpretations take both Socrates and Thrasymachus to be arguing rather badly in Republic I. I rather lean toward thinking they’re both arguing fairly well; but my present concern is with Thrasymachus, not with Socrates.

There are three possibilities here. One is that Plato deliberately made Thrasymachus’s position inconsistent on this point. But in that case it’s somewhat puzzling that Plato never has Socrates pick up on and exploit the inconsistency, but instead has him criticise the view on other grounds. A second possibility is that Plato himself simply failed to notice this glaring inconsistency; but that hardly seems the way to bet. The third possibility, and the one I shall defend as actual, is that, whatever its other deficiencies, Thrasymachus’s position is actually not inconsistent on the present point at all. (Well, a fourth possibility is that Thrasymachus’s inconsistency is innocent: he is simply switching from defining the concept of justice to using the terms ‘just’ and ‘unjust’ as they are conventionally used, meaning by them something like ‘what is conventionally considered just’ and ‘what is conventionally considered unjust’. But I take it that Thrasymachus’s definition is intended to capture how the terms are actually used; Thrasymachus is not arguing for a change in the reference of the terms, but is rather trying to lay bare their implicit sense.)

We can see a clue to the resolution of the apparent inconsistency by noticing how freely Thrasymachus conjoins P1 with the thesis that justice is another’s good (call this P5):

\[1\]

I owe this objection to Elizabeth Brake.
• **P5. If S acts justly, then S acts in the interest of someone other than S**

(My reasons for reading P5 as a conditional rather than a biconditional will become clear shortly.) P5 seems incompatible with P1, since in one case – the case where S is himself one of the stronger/rulers – S’s acting justly will count as self-regarding by P1 but as other-regarding by P5. Yet Thrasymachus evidently sees no conflict between them, since he happily equates justice as the advantage of the stronger/rulers with justice as the advantage of another, while contrasting justice as the advantage of the stronger/rulers with injustice as one’s own advantage:

> Justice and the just are really the good of another, the advantage of the stronger and the ruler. (343c)
> The just is the advantage of the stronger, while the unjust is the profit and advantage of oneself. (344c)

By virtually identifying the advantage of the stronger/rulers with the advantage of another, Thrasymachus seems to assume that the stronger/rulers are always someone *other* than the person whose justice or injustice is being assessed.

Why might he do this? One reason could be that Thrasymachus might regard the notions of justice and injustice as somehow applicable only to the conduct of the weaker/ruled;\(^1\) but on the contrary, we have seen that he is perfectly willing to describe the stronger/ruled as acting unjustly. The explanation must therefore lie elsewhere.

I wish to suggest a different reason: the apparent inconsistency between P1 on the one hand and P4 and P5 on the other is to be explained by Thrasymachus’s having a *relational* conception of the notion of stronger/ruler. To act in the interest of the stronger/ruler, in the sense intended by P1, is to act in the interest of someone *stronger-than-one-self*, of a *ruler-over-one-self*. The cause of our confusion about Thrasymachus’s meaning, on my interpretation, is the mistaken assumption that his phrases ‘stronger’ and ‘ruler’ always refer to the same person or group of people regardless of context, whereas in fact to be the stronger/ruler is always to be the stronger/ruler in relation to somebody (specifically, somebody *else*). And this would explain why Thrasymachus uses the comparative – the *stronger* rather than either the *strong* or the *strongest*.

The superficial similarity between Thrasymachus’s view and that of a class theorist like, say, Karl Marx is thus misleading. Marx, at least on a common interpretation, regards justice as determined by the interests of the ruling class,

---

\(^1\)Barker (1906) tries to help Thrasymachus out by explaining that ‘*for everybody other than the ruler*, justice may really and in truth be defined ... as “another’s good”.’ (95; emphasis added) Irwin (1995) likewise suggests that Thrasymachus’s definition ‘is concerned with behaviour by subjects, not behaviour by rulers’ (175); and Reeve (2006) offers a similar maneuver. (18) But of course Thrasymachus himself makes no such explicit restriction.
so that workers who respect capitalist property claims are behaving ‘justly,’
while capitalists who expropriate surplus value from the workers are likewise
behaving ‘justly.’ In short, actions that serve the interests of the capitalist class
are just (within capitalist society) regardless of whether the agents are
bourgeois or proletarian. It can be tempting to read Thrasymachus the same
way; but as we have seen, such a reading won’t work – since Thrasymachus
denies that the self-serving actions of the rulers count as just.

Thrasymachus is not identifying justice, then, with behaviour that serves the
interests of a particular identifiable ruling class regardless of whose behaviour
it is; rather, Thrasymachus’s example of benefiting an identifiable ruling class
is intended as a specific instance of benefiting the stronger, not as a
redescription of it. Indeed Thrasymachus infers the broader thesis from the
narrower one:

This, then, sir, is what I say the just is, alike in all cities – the
advantage of the established rule. And since the latter is what
exercises power, it follows for anyone who reasons correctly
[symbainei toi orthos logizomenoi] that the just is the same thing
everywhere – the advantage of the stronger. (338e-339a)

I take ‘the same thing everywhere’ (pantakhou ... to auto), i.e., in all
contexts, to be broader than ‘alike in all cities’ (en hapasais tais poleisin
tauton), which specifies a political context; I likewise take ‘the stronger’ (tou
ekreittonos) to be broader than ‘the established rule’ (tis kathseukias arkhês).
The general thesis that justice involves deference to the interests of a superior
is thus being inferred from the specific relation between what is called just
within a city and the ruling class within that city, rather than being a theory of
that relation exclusively.

Hence, on my interpretation, when a subject acts to benefit the ruler, he acts
justly, by putting a superior’s interests before his own; but when the ruler acts
in his own interest, he acts unjustly, since he pursues his own interests and
defers to no superior. The ruler could act justly only by submitting to some still
further ruler over himself.

By analogy, consider a possible definition of filial piety as reverence for the
parents. One might make the mistake of thinking that who counts as ‘the
parents’ is independent of context, so that there is some specific group of
persons called ‘the parents,’ reverence toward whom counts as filial piety.
And that would in turn imply that these persons, ‘the parents,’ would manifest
filial piety by expressing reverence for themselves. But in fact filial piety
means reverence for one’s own parents, so that which people one must show
reverence toward shifts depending on who one is. Athena can express filial
piety by revering Zeus, but Zeus can express filial piety only by revering
somebody other than Zeus. Analogously, then, on the interpretation I am
suggesting, I can express Thrasymachean justice by acting in the interest of my
rulers, but they can express Thrasymachean justice only by acting in the
interest of some yet further rulers.
The source of the ambiguity is the fact that P1, Thrasymachus’s definition of justice, is ambiguous between P6 and P7:

- **P6.** S acts justly iff S acts in the interest of the stronger/rulers, i.e., the dominant party in S’s political community.
- **P7.** S acts justly iff S acts in the interest of the stronger/rulers, i.e., those who are stronger than, or rulers over, S.

When P1 is interpreted as P6, then it entails P2 and P3, and is inconsistent with P4; but when P1 is interpreted instead as P7, then while it still entails P2, it no longer entails P3, and is now consistent with P4 (so long as ‘stronger/rulers’ in P4 is still understood non-contextually). Since Thrasymachus endorses P2 and P4 (so understood) while rejecting P3, it makes better sense to interpret Thrasymachus as intending P1 in the sense of P7 rather than in the sense of P6.

**Expertise and Mutual Deference**

Thrasymachus famously holds that the stronger/rulers count as stronger/rulers only when they are acting in their own interest; call this the Expertise Thesis:

*Do you suppose that I call the person committing an error stronger at the time when he errs? ... The ruler, insofar as he is a ruler, never errs – and, not erring, he decrees what is best for himself.* (340c-341a)

Introducing the Expertise Thesis is of course a fair enough move for Thrasymachus to make when debating Socrates, since Socrates himself believes something similar (if not so clearly time-indexed), namely that only those with wisdom count as genuine rulers (Plato, *Minos* 317a-318b; Xenophon, *Mem.* III.9.10-11; cf. Plato, *Euthyd.* 292a-d) and that only those who understand their true interests count as having genuine power (Plato, *Gorgias* 466b-470b).

Now Thrasymachus’s definition of justice, once it is understood as P7, turns out to have a non-accidental connection to the Expertise Thesis. For if a ruler were to rule in his subjects’ interests, as Socrates advises, then by subordinating his own interest to theirs the ruler would, by Thrasymachus’s lights, be deferring to his subjects’ authority, treating their interests as authoritative for him, and thus setting them up as something like rulers and authorities over him; no wonder, then, that rulers count as rulers only when they do not defer to their subjects’ interests.¹

¹Weiss (2007) comes close to my interpretation when she asks whether a ruler who obeys his own laws will be acting justly (since his laws are supposedly passed in his interest), and
Yet matters are a bit more complicated. For when the Expertise Thesis is conjoined with P7, it yields P8:

- **P8. S acts justly only if S acts in the interest of those who are acting in their own interest.**

  Suppose, then, that Arthur is Morgana’s nominal ruler, but defers to Morgana by subordinating his interests to hers. Then there are two possibilities. One is that Morgana responds by taking advantage of Arthur’s subordination for her own benefit; in that case it seems that by Thrasymachus’s lights it is now Morgana who is ruler over Arthur, rather than vice versa.

  But the other possibility is that while Arthur defers to Morgana’s interests, Morgana takes no advantage of this but instead defers in turn to Arthur’s interests. The possibility of this sort of mutual deference, as it happens, is acknowledged by Thrasymachus in his very first intervention, when he asks Socrates and Polemarchus: ‘Why do you act like simpletons by yielding to one another?’ (336b-c) If Arthur is deferring to Morgana’s interests then (for so long as he does so) he is acting like a simpleton and has clearly lost his title to be ruler/stronger; but if Morgana is deferring in turn to Arthur’s interests then she has no claim to the title of ruler/stronger either. And if neither is the ruler/stronger, then neither, in deferring to the other’s interests, is deferring to the advantage of a ruler/stronger – and so *neither is behaving justly.*

  Since they are not pursuing their own interests, presumably neither is behaving *un*justly either; but that by itself is no problem, since justice and injustice can be regarded as contraries rather than contradictories. But if pursuing another’s good were *sufficient* for being just, then we would have a contradiction; Arthur and Morgana would count as just by the deferring-to-another criterion, yet also as nonjust (albeit not unjust) by P8’s deferring-to-a-non-deferrer criterion. Thrasymachus never explicitly says, however, that deferring to another is sufficient for being just, so on grounds of charity we should decline to saddle him with such a claim; and that is why I made P5 a conditional rather than a biconditional. And since Thrasymachus thinks only a foolish few would be such simpletons as to respond to deference with like deference, deferring to another will *ordinarily,* even though not exceptionlessly, involve deferring to a non-deferrer; thus Thrasymachus’s near-equation of P1 with P5 will be justified despite one’s being a biconditional and the other not.

concludes that ‘at those times such a man will have ceased to be a ruler,’ since ‘the true ruler operates above the law and outside justice.’ (96) But although it will presumably in many cases be in the interests of the ruler to break the laws he requires his subjects to obey, it will not *always* be so (a ruler might wish to impose on his subjects a law against unduly risky or suicidal behaviour, for example, in order to be assured a healthy work force to tax and conscript; but it doesn’t follow that the ruler would benefit by breaking this law in his own case); hence Weiss’s law-centered way of drawing the distinction seems finally unpromising.
The Problem of Mutual Justice

I’ve argued that Thrasymachus’s position is not initially inconsistent. Later in the discussion, however, Thrasymachus makes a fatal concession when Socrates raises the question (351c-352a) of whether the stronger/rulers who are inflicting injustice on others will need to act justly toward one another in order to maintain effective cooperation. Socrates’ main goal here, which he achieves, is to get Thrasymachus to concede that justice can be in the interest of those who act justly; but what easily goes unnoticed is Thrasymachus’s logically more basic concession that there can even be such a thing as people exercising justice toward one another (be this in their interest or not).

I don’t think the significance of this concession goes unnoticed by Plato; I suspect it was not by accident that Plato initially has Socrates ask Thrasymachus whether the stronger/rulers couldn’t cooperate better with one another by not being unjust toward one another (rather than asking in positive mode whether the stronger/rulers couldn’t cooperate better with one another by being just toward one another).

Does it seem to you that a city, or an army, or robbers or thieves, or any other group of people joined together in some common unjust endeavour, would be able to accomplish anything if they were unjust toward one another?...What if they were not unjust toward one another? Would they not accomplish more? (351c-d)

This way of wording the question is perhaps intended to lure Thrasymachus into a more crucial concession than he realises, by making the looming contradiction less apparent.

But Socrates soon slides from talk of mutual refraining-from-injustice to talk of mutual justice, and Thrasymachus raises no protest:

For injustice, Thrasymachus, brings about factions and hatred and conflicts toward one another, while justice brings about unity of purpose and friendship [viz., toward one another]. – Let it be so, said he [= Thrasymachus], in order not to disagree with you. (351d)

Once Thrasymachus has granted, however half-heartedly, that it’s possible for people to act justly toward one another (en allēlois ... dikaiosumē), the asymmetry required by P7 can no longer be consistently retained. And this nicely prepares the way for Glaucon’s theory in Book II, since Glaucon’s theory is in the spirit of Thrasymachus but without asymmetry. (Glaucn also replaces the simple question of whether or not justice is beneficial to the just with the comparative question, beneficial relative to which alternative?)

But should Thrasymachus have made this concession? If justice involves sacrificing one’s own interests to those of others, while injustice (presumably – though Thrasymachus never explicitly defines injustice) involves sacrificing others’ interests to one’s own, then mutually beneficial cooperation would
logically have to count as *neither* just nor unjust. Thus Thrasymachus could consistently grant the benefits of mutual *non-injustice* without thereby granting so much as the *possibility* of mutual *justice*. If, then, Thrasymachus had resisted Socrates’ slide from non-injustice to justice, could he have held on to his theory without inconsistency?

Technically, perhaps. But the slide from mutual non-injustice to mutual justice goes by so easily precisely because it is so natural. Refraining from theft, murder, trickery, *et c.*, so obviously counts as justice (and not merely as refraining from injustice) when practised by the ruled toward the rulers that it is hard to justify denying its status as justice when practised by the rulers toward one another. If Thrasymachus were to deny its status as justice merely because those who practice it are benefiting from it, his position, while consistent, would be blatantly question-begging.

**Conclusion**

Thrasymachus’s theory of justice, then, is in the end not defensible. But it is not subject to the glaring inconsistency that has traditionally seemed to beset it when Thrasymachus’s relational, contextual use of terms like ‘stronger’ and ‘ruler’ is erroneously interpreted non-relationally and non-contextually. The apparent inconsistency between justice as benefiting the stronger/ruler and justice as benefiting the other – criteria that seem to generate opposite results whenever one is oneself the stronger/ruler – is dissolved once we realise that one *never is* oneself the stronger/ruler in Thrasymachus’s sense, since “stronger/ruler” is defined as stronger than, or ruler over, oneself. Once we see how Thrasymachus avoids this inconsistency, we gain a better understanding of what Thrasymachus’s view of justice actually is.

**Bibliography**


In *Physics IV*, Aristotle develops an account of time as a kind of number. I will argue that his account of time is built upon the idea that time has two constituents, one of which functions in the being of time as form functions in the being of a familiar particular, and the other of which functions in the being of time as matter functions in the being of a familiar particular.

I will begin with an overview of Aristotle’s discussion of time. He argues that time is not simply motion; for motion only occurs where there is some changing thing, or within a changing thing. Time, on the other hand, is omnipresent. Moreover, the same movement cannot occur in two distinct places, while it can be one and the same time in two distinct places. And motion can occur more or less quickly, while time cannot; for time is itself the measure of quickness. We would have to posit meta-times *ad infinitum* in order to make sense of time as having speed (218b10-20). Although time is not motion, Aristotle argues that neither is it independent of motion. For it is only through an awareness of motion that we become aware of a lapse of time (218b21-219a2). He concludes that time is an attribute of motion: it is dependent upon but not reducible to motion.

Aristotle next begins to craft his analogy between time, motion, and magnitude. He writes that ‘what is moved is moved from something to something, and all magnitude is continuous…. Because the magnitude is continuous, the movement too is continuous, and if the movement, then the time’ (219a10-14). The continuity of motion arises out of the continuity of magnitude, and in a parallel fashion, the continuity of time arises out of the continuity of motion.

His analogy also involves the concepts of before and after at each of its three tiers. The only way to determine a magnitude is to divide a line at two points that differ in their relative positions. In terms of this difference in their relative positions, Aristotle refers to one such point as before and to the other as after. He writes:

---

1All references to Aristotle’s work are taken from Barnes (1984).
The distinction of before and after holds primarily, then, in place; and there in virtue of relative position. Since then before and after hold in magnitude, they must hold also in movement, these corresponding to those. But also in time the distinction of before and after must hold; for time and movement always correspond with each other. (219a15-20)

Like continuity, the before and after are manifested at each tier of the analogy, and the before and after in time arise out of the before and after in motion, as the before and after in motion arise out of the before and after in magnitude. G.E.L. Owen and Edward Hussey agree that Aristotle’s analogy grounds a relationship of structural dependency between magnitude, motion, and time. Owen writes, Aristotle ‘wants to show that spatial order is conceptually basic to the rest, that by starting from this we can explain the order of movement, and at another step the order of time, without circularity’ (1986, 313). And Hussey writes:

Both for the pair ‘magnitude-change’ and for the pair ‘change-time’, Aristotle’s claim is twofold: (i) there are relationships of priority…in virtue of which the second member of each pair is dependent upon the first; (ii) as a result of this dependence, the second member of each pair has an internal structure analogous to the first, and derived from it. (1983, 142).

Aristotle’s analogy thus grounds a dependency between magnitude, motion, and time in terms of their continuity and in terms of their manifestations of the before and after. But Aristotle also needs to maintain the sui generis reality of each tier, since he holds that time cannot simply be reduced to motion.

A Constituent Model for Time, Motion, and Magnitude

I propose to understand the ontology of each of magnitude, motion, and time in Aristotle’s analogy as modeled after his ontology of a familiar particular. For Aristotle, familiar particulars are constituted by the union of a form with matter. But a familiar particular’s proximate matter is itself a composite of more primitive matter and a lower-level universal; and the proximate matter of this familiar particular’s proximate matter is also a composite of more primitive matter and a lower-level universal; and so on. This hylomorphic analysis cannot be carried out ad infinitum; rather, it will lead us to an ultimate subject of predication. Aristotle writes that ‘the ultimate substratum is of itself neither a particular thing nor of a particular quantity nor otherwise positively characterized; nor yet negatively, for negations also will belong to it only by accident’ (1029*24-26). Traditionally, this passage has
been interpreted to mean that the ultimate substratum has no actual essence.¹ A
full analysis of a familiar particular’s matter thus requires an appeal to
essenceless prime matter, which exists per se merely potentially.

The other constituent of a familiar particular, its form, is what Aristotle
ultimately comes to recognize as the cause of its being: “Therefore what we
seek is the cause, i.e. the form, by reason of which the matter is some definite
thing; and this is the substance of the thing” (1041b7-8). Its form is the source
of a familiar particular’s actuality: “the proximate matter and the form are one
and the same thing, the one potentially, the other actually” (1045b18-19). Its
form, as the cause of its being and its actuality, endows a familiar particular
with its essence. Form (unlike matter) is simple; it has no constituents.

I will interpret Aristotle as arguing that there are two necessary and jointly
sufficient conditions for the existence of each of magnitude, motion, and time:
1) the existence of a constituent that functions analogously to the way in which
form functions in the being of a familiar particular (I will refer to this
constituent as the formal constituent) and 2) the existence of a constituent that
functions analogously to the way in which matter functions in the being of a
familiar particular (I will refer to this constituent as the material constituent).

Aristotle defines time as ‘number of motion in respect of “before” and
“after”’ (219b1). How are we to relate his talk of the now to this definition? He
writes, ‘The “now” in one sense is the same, in another it is not the same. In so
far as it is in succession…it is different…but its substratum is the same’
(219b13-15). How does its being in succession render the now always
different? He writes, ‘number…according as it occurs before or after is always
different, for the “nows” are different’ (220b9-10). Two nows differ insofar as
one picks out the before and one picks out the after mentioned in the definition
of time.

Aristotle goes on to reiterate that ‘the “now” as substratum remains the
same (for it is what is before and after in movement)’ (219b26-27). He is
claiming that in one sense the now remains the same, and then giving his
reason for this claim: because two nows number the before and after of the
same underlying movement. Hussey, who argues that Aristotle sees the now as
both the source of the continuity of a time (insofar as it is ‘the permanent
present’) and the differentiating aspect of a time (insofar as it is an
‘unrepeatable instant’), interprets this passage as asserting that the-now-
insofar-as-it-remains-the-same (according to his interpretation, ‘the permanent
present’) is identical with the before and after in movement (1983, 153-157).
But if ‘the permanent present’ is identical with the before and after in
movement, then ‘the permanent present’ must be just as confined to particular
movements as are the before and after in movement. And then time cannot be
everywhere the same; there will be a different time for every different
movement that is going on (1983, 155). Hussey flees from this result by
moving toward an interpretation of time as ‘not a particular change qua
measurable, but as an abstraction from all changes, an abstract change-qua-

¹Some commentators have challenged this conclusion.
measurable’ (1983, 155). This move, however, contradicts another of Aristotle’s claims: ‘Time, then, is [number in the sense of] what is counted, not that with which we count’ (219b7-8). If time is abstract number of motion, it is number in the sense of that with which we count.

By locating the source of the sameness of two nows in their role in numbering one and the same movement (rather than in their serving successively as the permanent present), I can interpret this passage in a way that avoids Hussey’s identification of ‘the permanent present’ with the before and after in movement. Aristotle’s claim is not that the-now-insofar-as-it-remains-the-same is identical to the before and after in movement; it is rather that two nows are in one sense the same because, taken together, they number the before and after of the same movement.

Analogue of my interpretation at the level of time apply at the other two levels of Aristotle’s analogy. Two coincidentals of a moving body, like two nows, are the same in one respect and different in another:

[T]here corresponds to the point the body which is carried along, and by which we are aware of the motion and of the before and after involved in it. This is an identical substratum...but it is different in definition—as the sophists assume that Coriscus’ being in the Lyceum is a different thing from Coriscus’ being in the market-place. (219b16-21)

Two coincidentals of a moving body (Coriscus-in-the-lyceum and Coriscus-in-the-market-place) are the same in that they are both coincidentals of the same underlying body (Coriscus). Yet they differ in that they predicate different accidents of Coriscus: ‘the body which is carried along is different, in so far as it is at one time here and at another there’ (219b21-22). They differ insofar as one coincidental is the before in the change and one is the after, just as the two nows differ in respect of whether they are before or after in a time.

Two points on the same line are, like two nows and two coincidentals of a moving body, in one respect the same and in another respect different:

The ‘now’ is the link of time... (for it connects past and future time), and it is a limit of time (for it is the beginning of the one and the end of the other). But this is not obvious as it is with the point, which is fixed. It divides potentially, and in so far as it is dividing the ‘now’ is always different, but in so far as it connects it is always the same, as it is with mathematical lines. (222a10-15)

Two points differ in respect of whether they occur before or after relative to one another. Indeed, Aristotle writes, ‘The distinction of before and after holds primarily...in place; and there in virtue of relative position’ (219a15-16). Yet two points are the same in that they both have their being through the division of the same mathematical line.
Now, two points jointly bear a relation to the line on which they are found that is analogous to that which form bears to matter. For it is in virtue of form that matter is determined to constitute a familiar particular of a certain essence, and it is in virtue of its essence that a familiar particular is different from or similar to other familiar particulars. Just so, a magnitude has its determinate length in virtue of its endpoints, and it is in virtue of its length that a magnitude is similar to or different from other magnitudes. The underlying line, in turn, serves as the material constituent of the magnitude. Indeterminate in itself, the line is delimited by points into a finite magnitude.

The analogy goes deeper. Aristotle writes: ‘What is continuous is divided ad infinitum, but there is no infinite in the direction of increase...since no sensible magnitude is infinite, it is impossible to exceed every definite magnitude’ (207b16-21). As a finitist, Aristotle thinks that a line, because it is infinite, can never be found as such in the actual world; it can only be found delimited by endpoints into a finite magnitude. Matter can be analogously analyzed (given the traditional doctrine of prime matter). For, again, the proximate matter of a familiar particular is itself a composite of prior matter and a universal, and this complexity continues all the way down to the level of prime matter. Full hylomorphic analysis of a familiar particular’s matter requires us to appeal to essenceless prime matter. But just as there could never be a line not delimited by points into a finite magnitude, so there could never be any prime matter not endowed with a determinate form of being. In and of itself, a line has merely potential existence, as does matter in its ultimate analysis (i.e., prime matter). Indeed, Aristotle explicitly makes this connection: ‘The infinite, then, exists in no other way, but in this way it does exist, potentially and by reduction. It exists...potentially as matter exists’ (206b13-15). These considerations support the conclusion that a line serves as the material constituent of a magnitude, while two points jointly serve as the formal constituent.

Two coincidentals, like two points, differ in respect of whether they are before or after in a movement, and they determine the kind of change that occurs in the way that two points delimit a magnitude of a certain length, and analogously to the way in which a form determines a familiar particular’s kind. Moreover, just as matter could never be found apart from any universal, nor a line in its full infinitude, so the single body underlying both coincidentals will always be united with some accident in the constitution of some coincidental. An underlying body, then, serves as the material constituent of a movement, and two coincidentals of that body’s movement jointly serve as the formal constituent.

The before-now and the after-now jointly determine a stretch of time as points determine a magnitude, coincidentals a movement, and form a parcel of matter; for time just is ‘what is bounded by the “now”’ (219a28). Indeed, Aristotle writes that ‘if the “now” were not different but one and the same, there would not have been time’ (218b27-28). Two nows, then, jointly serve as the formal constituent of time. Number of motion serves as the material constituent, since it is delimited by a before-now and an after-now to constitute
a time. And just as there is nothing human about matter apart from the human form, so there is nothing temporal about number of motion apart from nows. Now are the source of the *sui generis* properties of time, as the human form is the source of the *sui generis* properties of a human being.

Just as matter and form are metaphysically prior to the entity they constitute, so number of motion and the now are metaphysically prior to time. But we should not therefore expect to find number of motion or the now apart from any time, just as we do not expect to find formless matter, or matterless form. Just as matter could never be found apart from form, nor the infinite mathematical line undetermined by endpoints, nor a body apart from any accidental, so number of motion could never be found apart from any nows. Conversely, the now can only be found along with number of motion in the constitution of some time; for Aristotle writes, ‘Clearly, too, if there were no time, there would be no “now”, and vice versa’ (220\(^b\)1-5).

Aristotle writes, ‘the “now” corresponds to the body which is carried along, as time corresponds to the motion. For it is by means of the body that is carried along that we become aware of the before and after in motion, and if we regard these as countable we get the “now”’ (219\(^b\)22-25). Julia Annas interprets this passage as claiming that ‘the now is related to the moving body as time is related to the motion of the moving body…the relation of time to the motion is that time is the number of motion. Correspondingly, the now must be the number of the moving body” (1975, 110). She concludes that this passage does not fit in with the rest of the *Physics* IV account of time, and is merely a less careful phrasing of a point which Aristotle goes on to state: ‘For the number of the locomotion is time, while the “now” corresponds to the moving body, and is like the unit of number’ (220\(^a\)3-5).

My model provides a way of interpreting the first passage which both renders it a coherent part of the *Physics* IV account and maintains its consistency with the second. Aristotle’s claim at 219\(^b\)22-25 is not that the now is the number of the moving body, as time is the number of motion. It is rather that the now corresponds to the body which is carried along because both serve as the formal constituent in the entity they constitute (time and motion, respectively), while time corresponds to motion because these are the constituted entities. On my reading, this passage supports the analogy articulated throughout the *Physics* IV account of time.

My interpretation also maintains the consistency of this passage with Aristotle’s subsequent claim that the now is like the unit of number. For if the point of the former passage is that the now serves as the formal constituent in the being of time, then there is a sense in which it is like the unit of number. As the unit of number determines what number we will ultimately end up with in counting (by telling us precisely what to count), so two nows together determine what time they constitute (by delimiting its duration). Furthermore, whereas Annas concludes that inconsistency lurks between Aristotle’s comparison of the now with a unit of number (which must have extension) and his claim that the now is durationless (1975, 111), my interpretation of the sense in which the now is like the unit of number does not require the now to
have temporal duration, and so renders these claims consistent: Two
durationless nows together determine a certain time by marking its endpoints,
as the unit of number determines what number counting will yield.

Two points jointly serve in the being of a magnitude, two coincidentals in
the being of a movement, and two nows in the being of a time as a constituent
that functions analogously to the way in which form functions in the being of a
familiar particular (by marking the before and after which determine a
magnitude, a movement, or a time, respectively). And a line serves in the being
of a magnitude, a body in the being of a movement, and number of motion in
the being of a time as a constituent that functions analogously to the way in
which matter functions in the being of a familiar particular (each is the source
of continuity at its level and is delimited by points, coincidentals, or nows into
a determinate entity, as form actualizes the potentiality of matter and
determines it into an entity with an essence).

Maintaining Dependency and Irreducibility

Each tier of Aristotle’s analogy must be *sui generis*, since he holds that time
is not merely reducible to motion. Still, Aristotle wants to maintain a
dependency running from magnitude to motion to time with respect to both the
continuity manifested at each tier and the manifestations of the before and after
at each tier. My model upholds both aspects of Aristotle’s view. Consider
number of motion, time’s material constituent. Number of motion could not
possibly exist apart from motion. Hence, the substratum of time could not be
without motion. Time is thus ontologically dependent upon motion. ower,
number of motion is not all there is to time; nows too are required for its
constitution. Time is thus not reducible to motion.

One might object, ‘Your location of the *sui generis* character of time in
nows does not comport with Aristotle’s elucidation of the dependency of the
before and after in time on the before and after in motion. For the before and
after in time and in motion are marked by nows and coincidentals, respectively;
thus, Aristotle is emphasizing the dependency of that which you have argued to
be *sui generis*.’ But given my model, this dependency does not destroy *sui
generis* character. Consider the human form, which cannot exist apart from
matter; it is thus ontologically dependent on matter. A human being’s
proximate matter is in turn constituted by some lower-level universal (e.g.,
fleshiness). It follows that the human form is ontologically dependent upon this
lower-level universal; but it does not follow from this that the human form is
not *sui generis*. Analogously, on my model nows are ontologically dependent
upon number of motion, as form is upon matter; and number of motion is in
turn ontologically dependent on coincidentals (since coincidentals serve as the
formal constituent of motion). It follows that nows are ontologically dependent
on coincidentals, but this dependency does not affect the *sui generis* reality of
nows any more than the ontological dependency of the human form on a lower-level universal affects its *sui generis* reality.

**Defending the Reality of Time**

The existence of a time is, on my model, nothing over and above the obtaining of a state of affairs in which two nows unite with number of motion, just as the being of a familiar particular is nothing over and above the obtaining of the state of affairs in which a form is instantiated by a parcel of matter. Does it follow that time is merely a heap of these constituents, and not a genuine unity of its own?

Michael Loux has proposed a response to this objection for the case of the familiar particular. Although the existence of, say, a human is nothing over and above the existence of its two constituents, form and matter, one of these constituents, the human form, is ‘equideterminate with the species’ (2005, 119). This form cannot be found apart from any human, and no human can be found apart from it. Thus, the human form is not a *lower-level* constituent of a human. And in order for a composite entity of a kind *K* to be reducible to its constituents, in order for it to lack *sui generis* reality, its constituents must both be lower-level constituents—they must be able to exist apart from any *K* (2005, 118).

This solution can also apply to time on my interpretation. Aristotle writes, ‘Clearly, too, if there were no time, there would be no “now”, and vice versa’ (220a1). Now, which serve as the formal constituent of a time, are equideterminate with time; they cannot be found apart from any time, nor any time apart from them. We can say that time is fully real in virtue of having a constituent equideterminate with itself.

**Answering the Arguments for the Unreality of Time**

At the beginning of *Physics* IV, Aristotle considers two arguments for the unreality of time (217b32-218a29). The first proceeds as follows: If a thing has parts, then, necessarily, it only exists if all or at least some of its parts exist. It seems that the past, present, and future are the parts of time. But the present is ultimately no more extensive than the now, and ‘the “now” is not a part; a part is a measure of the whole, which must be made up of parts. Time, on the other hand, is not held to be made up of “nows”’ (218a6-8). The extensionless now is too fleeting to be a part of time. This leaves only the past and the future; however, the past has been but is not, and the future will be but is not. No part of time, then, exists. And ‘[o]ne would naturally suppose that what is made up of things which do not exist could have no share in reality’ (218a4-5).

Aristotle never explicitly responds to this argument. Owen argues that although Aristotle’s account upholds the reality of time, since it shows that ‘[j]ust as there cannot be points without lengths that they divide and join, so
there cannot be temporal points—“nows”—without...the stretches of time past and future that meet in them’ (1986, 310), the account still falls prey to this argument, since ‘at any real present the measurable stretches of time are behind or in front of us’ (1986, 314).

On my model a response to this argument becomes clear. Although it is true that the mereological parts of time, the past and the future, do not exist, it does not follow that time does not exist. For time also has metaphysical parts, or constituents: number of motion and nows. Although the now cannot serve as a mereological part of time because it lacks extension, it can serve as a constituent of time, because constituents function differently than mereological parts. Consider the spatial case of a human being: the human form could not possibly serve as a mereological part of a man, since it has no spatial extension; but it does serve as a constituent of a man. Likewise in the temporal case, whereas a mereological part of time must be temporally extended, a constituent need not meet this requirement. Thus, on my model, time can exist even though none of its mereological parts exist, because its constituents exist. Indeed, Aristotle thinks a familiar particular can survive the loss of its mereological parts (e.g., a human’s hand), but not the loss of its constituents (e.g., a human’s form) (Loux, 2006). If Aristotle thinks that time too has constituents, he could respond similarly to this argument against the existence of time.

The question underlying the second argument for the unreality of time which Aristotle considers regards ‘the “now” which seems to bound the past and the future—does it always remain one and the same or is it always other and other?’ (218a9-10). If the former, then ‘things which happened ten thousand years ago would be simultaneous with what has happened to-day, and nothing would be before or after anything else’ (218a27-29). If the latter, then ‘the prior “now” must always have ceased to be’ (218a16). But there is no time at which it could have ceased to be: not while it existed, because then it would have simultaneously both existed and not existed; and not during a now other than itself, because there are infinitely many nows between any two given nows, as there are infinitely many points between any two points. Thus if the now were to cease to be during a now other than itself, ‘it would exist simultaneously with the innumerable “nows” between the two—which is impossible’ (218a20-21).

Faced with this dilemma, Aristotle answers the question of whether the now always remains the same or is always other and other by saying, ‘[t]he “now” in one sense is the same, in another it is not the same’ (219b13). My model, according to which two nows together stand to one and the same number of motion as the formal to the material constituent of time, provides metaphysical support for Aristotle’s claim. As a familiar particular’s matter and form are numerically one (though different in definition), so two nows are numerically one with the same substratum, number of motion (though different from it in definition). It follows that these nows themselves are numerically one, since they jointly serve as a constituent of numerically one time. But they can still
differ from each other in definition—in respect of whether they are before or after—as indeed they must in order to serve as the formal constituent that determines number of motion into a time. Thus, if two nows jointly serve as the formal constituent and number of motion as the material constituent of a time, it follows that these nows are numerically the same, yet different in definition.

The Soul Argument

Toward the end of *Physics* IV, Aristotle argues that time could not exist without soul:

> Whether if soul did not exist time would exist or not, is a question that may fairly be asked; for if there cannot be someone to count there cannot be anything that can be counted either, so that evidently there cannot be number; for number is either what has been, or what can be, counted. But if nothing but soul, or in soul reason, is qualified to count, it is impossible for there to be time unless there is soul. (223a22-28)

In this argument, Aristotle is making distinctions which have no modal implications. For the argument may be thwarted from the very beginning by Aristotle’s principle of instantiation: It is impossible that there be no soul, and therefore it is impossible that there be no one to count. And if it is impossible that there be no one to count, then it is impossible that there be nothing that can be counted. But then it is impossible for number to fail to exist, because ‘number is either what has been, or what can be, counted’ (223a24-25). Because soul could not have failed to exist, it follows that time could not have failed to exist. Why, then, did Aristotle bother with such a refined argument?

Consider my model, according to which Aristotle’s distinctions between time and number of motion (time’s material constituent) and between time and the now (time’s formal constituent) are not modal distinctions at all. They are not modal distinctions because a now could not possibly be found apart from time, and neither could number of motion (just as a form could not possibly be found except in the constitution of some composite entity, and neither could matter). The distinctions between number of motion and time and the now and time are, on my model, distinctions between time and its constituents—they are important distinctions indeed, but they have no modal implications. Given that Aristotle’s account of time is to be understood in this way, we would expect him to write as though it were important to make some distinctions which have no modal implications, simply because such distinctions are of central importance in elucidating a whole’s constituent structure.

Moreover, my model maintains the coherency of this argument. For one of the constituents of time is number of motion, which could not exist if number did not exist. Since number in turn could not exist if there were no soul, it indeed follows that number of motion, and hence time (since number of motion
is a constituent of time), could not exist if there were no soul. Given my model, this argument discusses an interesting implication of the metaphysical structure of time.

**Conclusion**

According to my interpretation, Aristotle is arguing that magnitude, motion, and time each have constituents which function in their being analogously to the way in which form and matter function in the being of a familiar particular. Understanding time to be constituted by number of motion and nows upholds both the dependency of time on motion and the *sui generis* character of time. This model of time connects Aristotle’s definition of time as ‘number of motion in respect of “before” and “after”’ (219b1) to his analogy between magnitude, motion, and time; for it finds both number of motion and the before-now and after-now to play constitutive roles in the being of time. This model also grounds the reality of time and furnishes responses to the arguments for the unreality of time which Aristotle considers. Furthermore, this model renders Aristotle’s argument that time could not exist without soul both coherent and interesting.

**Bibliography**

CHAPTER SIX

Is Machiavelli a Tragic Philosopher?

Ursula Niklas Peterson

The intellect of man is forced to choose
Perfection of the life, or of the work
W. B. Yeats, ‘The Choice’

The many questions asked about Machiavelli I want to add this one: Is Machiavelli a tragic thinker? I propose to approach that question from the historical perspective and to consider the role Machiavelli’s thought plays in the philosophy’s turn to the tragic. My paper takes as a point of departure the antagonistic historical relationship between tragedy and philosophy. The tragic, the vision of life which insists on the permanent possibility of moral conflict, was first articulated in the poetry of Homer and Attic tragedians. The rivalry between tragedy and philosophy was old already in the time of Socrates, as Heraclitus wanted to have the poets flayed with sticks, but only the emergence of Platonic philosophy spelled the demise of tragedy. The tragic was rejected by Plato whose philosophy offered a vision of the timeless cosmic harmony that supplies the purpose and the principle of order to human life. A well known broad historical narrative of the tension between tragedy and philosophy, outlined by Nietzsche (2000), tells the story of the birth of tragedy, its suppression by philosophy, and culminates in the coming of another tragic age heralded by Nietzsche who accords to himself the role of the first tragic philosopher (1985). The tragic philosopher, according to Nietzsche, grasps the nature of the tragic as the clash of reason and life, or the Apolline clarity and the Dionysian disorder, and announces the emergence of a new tragic philosophy.

Machiavelli certainly does not fit Nietzsche’s notion of the tragic philosopher. He neither refers to tragedy in his writings, except for a mention in his correspondence, nor does he model his philosophy on the wisdom of the ancient poets. Yet it is Machiavelli rather than Nietzsche who deserves the title of the first tragic philosopher. I argue that Machiavelli turns to the tragic by breaking away from the Platonic tradition. He is the first philosopher to reject the Platonic understanding of human life in the light of the timeless pattern of order and harmony. His vision of the world can be compared to the Platonic cave, but without its antipode of the transcendent realm. In Machiavelli’s philosophy there is no transcendent reality that would supply the telos to human life and no principle of cosmic order to which the soul should attune. He is the first thinker who reverses the Platonic hierarchy and elevates the active life over the contemplative life. According to Machiavelli, the striving for goodness in public life is endangered by the unstable and uncontrollable
circumstances of action, by limitations of knowledge, and by attachment to the particular and loyalty to what is one’s own. Machiavelli’s claim that the possibility of living a moral life is endangered by the very nature of what is worth living for sets him against the Platonic tradition and brings him close to the tragic view of the world.

Yet to read Machiavelli as a tragic philosopher is not to impute to him an attempt to revive what irrevocably belongs to the past. Such attempts are futile. Machiavelli’s philosophy develops the tragic theme in a radically new way. His world of finitude, contingency, and temporality is alien to the ancients, and in that world the tragic aspect of human life acquires a new meaning.

The ancient conflict between tragedy and philosophy plays itself out in the broader historical context of the transformation of Greek culture during the classical period. Plato’s attack on the authority of the poets, who, according to Herodotus (2007, IL.53), gave Greeks their gods and described the divine genealogy and attributes, becomes an attack on the traditional culture sustained by poetry. Plato condemns Attic tragedians and Homer, ‘the most poetic and first of the tragic poets,’ and seeks to undermine their traditional moral and educational authority by questioning their wisdom concerning what is worth living for and how to manage human affairs. Poetry is unable to grasp truth, Plato has Socrates argue in the Republic X (1961), because it speaks of particulars and merely imitates their appearances. Unlike philosophy, which views particulars in the light of universals, poetic creation fails to penetrate the surface of particulars and reach their universal natures.

Attachment to the particular and loyalty to what is one’s own are portrayed in tragic poetry as constitutive of the good life. Achilles’s love for Patroclus demands that he revenge his friend’s death even at the cost of his own life; Antigone sacrifices her life to fulfill her obligation to her dead brother. Attachments to individual persons make human beings vulnerable to the pain of irreversible loss and, more importantly, lead to a clash of moral principles. Agamemnon’s moral dilemma arises from the choice he has to make between his obligations as father and general. Antigone and Creon embody the conflict between the unwritten moral laws of family obligation and the political law of the city. In the Platonic philosophy, on the other hand, the love of the individual is subordinated to the love of the general. The desire for particular persons, Diotima teaches in the Symposium, occupies the lowest rung in the ladder of ascension to the proper love of the beautiful. The philosophical subordination of the particular to the general obliterates loyalty to what is one’s own; the philosophical positing of the general as the proper object of human striving and desire secures moral harmony and prevents the emergence of moral dilemmas that are the dominant themes of tragic poetry.

Machiavelli’s vision of the life worth living is dominated by the love of the particular. In the comedy Mandragola, which is a mirror held to private life, Machiavelli portrays characters who assert their passions and compromise their attachment to moral principle in the pursuit of desire. In public life, he asserts, one’s native country becomes the highest object of love. ‘A man is under no greater obligation,’ Machiavelli states (1961), ‘than to his country (patria). . . .
In fact he who shows himself by thought and deed an enemy of his country deserves the name of parricide, even if he has legitimate grievance. For if it is an evil deed to strike one’s father and mother for any reason, it necessarily follows that it is even more criminal to savage one’s country.’ Machiavelli is clearly and painfully aware that to affirm the loyalty to one’s own country it to risk the possibility of a moral conflict, and poignantly states this dilemma in a letter to his friend Vettori: ‘I love my country (patria) more than my soul’ (1996a).

Machiavelli’s view on what makes human life worth living is rooted in his naturalistic vision of the world. Machiavelli’s world is devoid of transcendental dimension and provides no source of order, meaning and value. Human affairs move within the cycle of recurrent transformations, and their circular paths never straighten in the direction of an unalterable and transcendent telos. In the mutable temporal world the cycle of change is prescribed by nature that commands things to move and sets limit to their life (1996b, III.1.1; 1988, V.1). Religions go through the cyclical changes, as do all human things, and one religion replaces another every few thousand years; in Machiavelli’s estimate, Christianity will last no longer than two millennia (1996b, II.1.4; 5.1). Although religion is divested of a supernatural aura, its human origin does not diminish its significance. Machiavelli accords to the founding of religion the highest place among human achievements since religion institutes moral order which is necessary for maintaining a civilization (1996b, I.11.1; 10.1). In his view, the morality Christianity introduced demonstrates its inferiority to the ancient pagan religion. He argues that the moral and political order sustained by ‘our religion’ weakens the individual and the state. Christianity glorifies humble and contemplative men rather than active men and encourages contempt for the human things, while the pagan religion glorifies striving for the greatness of spirit and the pursuit of glory. The idea of the good life professed by Christianity directs human gaze toward the other-worldly and away from the honors this world bestows upon action, especially upon political action in service of one’s native country (1996b, II.2.2). Machiavelli’s criticism of Christianity, however, does not stem from an anti-religious sentiment, but it belongs to his larger philosophical project of reversing the classical hierarchy of lives which places active life beneath the dignity of contemplative life. His criticism of Christianity and his praise of the pagan religion are motivated by the philosophical move that elevates the man of action above the contemplative man.

Machiavelli is not the first philosopher to note that the life of action involves moral transgression. What is new is his insistence that the man of action is under the necessity of not being good. This idea is contradictory to the tradition of classical philosophy which links moral transgression to the weakness of character, the all-too human failure rationally to control passions, or the deplorable ignorance of the truly good. The traditional indictment of human frailty, in spite of its harshness, is essentially optimistic because it is based on the assumption that the rational and harmonious order of cosmos supplies the pattern of goodness and provides direction to human moral
striving. According to Machiavelli’s tragic view, human beings violate moral principle by necessity (1998, XV), hence the inevitability of moral transgression is written into the human condition. Machiavelli is careful to distinguish his notion of necessity from the workings of a supra-worldly power that leave no room for free will. He rejects the view held by some of his contemporaries that worldly things are governed by God and Fortune, the pagan deity transformed into a servant of Providence whose rulings cannot be opposed or corrected with prudent actions (1998, XXV). Machiavelli’s naturalistic understanding of the world (Parel 1992) does not allow for transcendent constraints on free will. He demotes the figure of fortune to the sphere of worldly things and compares her to the river and to the woman, both strikingly naturalistic similes which he invests with the associations of threat, struggle, and conquest (1998, XXV). Machiavelli declares that fortune controls a half of our actions and leaves the other half for us to govern, and that we are free to oppose her and to enlarge the domain of autonomous human action. His Fortuna symbolizes the force of contingency that cannot be entirely eliminated from human life. ‘I indeed affirm it anew to be very true,’ he states, ‘that men can second fortune, ... can weave its warp but not break it’ (1996b, II.29.3).

The excellence of human action in the world of contingency and change, or Machiavellian virtu, is measured by the agent’s ability to deal with the vicissitudes of life symbolized by fortune. Human beings are vulnerable to reversals of fortune as they find themselves in the flux of events that are neither fully predictable nor controllable, and have to master the unstable circumstances of action which are not of their making. The excellence of the man of action lies in his ability to combat the vicissitudes of fortune by alternating between virtue and vice and by mastering the crafts of deception and violence. Thus the prince, Machiavelli famously (or infamously) said, must learn how not to be good and to use that knowledge according to necessity (1998, XV).

It is tempting to read Machiavelli as a teacher of evil (Strauss 1958). However, he emphatically asserts the validity of moral values and presents a catalogue of virtues and their corresponding vices for which everybody, and especially those who occupy prominent political positions, should be praised or blamed. Machiavelli’s discussions of specific virtues and vices, such as liberality and generosity, mercy and cruelty, trustworthiness and deception, open with a recurring theme: it is desirable and praiseworthy to have these virtues, but the person who enters political life must be prepared to compromise his attachment to moral principle. ‘Goodness is not enough,’ Machiavelli states, to control the circumstances of politics, and the life of action requires virtuosity in adapting oneself to the changing conditions (1996b, III.30.1). Virtu, or excellence in action, inevitably clashes with traditional virtue. The insufficiency of goodness has its source in the necessity to deal with fortune, and it is the very nature of human action that makes moral transgression inescapable.

Rather than a teacher of evil, Machiavelli is a tragic thinker who recognizes that the presence of moral conflict in human life does not stem from the
ignorance or weakness of character of the agents, but that the necessity of not being good is written into the human condition (le condizioni umane). According to Machiavelli, the lack of harmony is not a deplorable aspect of life with which we must contend. To the contrary, strife and disunion are sources of greatness. The conflict between the nobles and the people, he argues, was the source of stability of the Roman republic (Pocock 1975). Rome perfected herself through accidental rather than planned events, and she developed into a perfect republic by creating the political space for the conflicts to play themselves out. From the Platonic point of view, conflict and disharmony which pervade political life are symptoms of corruption of the perfect order and must be eliminated from the ideal society. Consequently, the Platonic vision of the best society has a peculiarly anti-political character (Wolin 2004). Against Plato, Machiavelli’s affirmation of political life is an affirmation of conflict, disunion, and strife as positive forces, and his recognition of the permanent possibility of a violation of moral principle in human life marks the philosophy’s turn to the tragic.

Bibliography

Modern and 19th Century
C H A P T E R  SEVEN

The Problems of Hobbes’s Concept of the Sovereign

Hsuan Huang

One of the questions asked by Hobbes in his *Leviathan* can be stated as follows: as human beings, what can we do to live well? This question resulted in a monstrous treatise, which left an indelible mark in the history of political philosophy. His answer to this question is suggested by the very title of his book. After analyzing the human nature, Hobbes concludes that the only way for us to attain ‘the felicity of this life’ is to establish a ‘great LEVIATHAN called a COMMONWEALTH, or STATE, (in Latin CIVIAS) which is but an artificial man.’ The person who bears the artificial person is called sovereign, who has ‘sovereign power,’ and everyone other than this person is his subject. The sovereign has ‘unlimited’ power, because all his subjects covenant with each other to stand out of his way and authorize him to represent them, so he can exercise the absolute liberty that he originally has and uses all the strength from his subjects, in order to protect them and maintain peace, which is a necessary condition for the subjects to own private property.

Gert points out that some philosophers seem to find Hobbes’s concept of human nature unsatisfactory, because ‘he does not put forward any view of the best life.’ While it is true that Hobbes does not think it is possible for us to live the best life, he develops his theory of living well. For Hobbes, a good life is a life in which one can keep one’s goods and enjoy them alone without being molested, and this cannot be achieved without the establishment of a commonwealth. The problems of his theory of living well, from my point of view, mainly arise from his concept of the sovereign. The sovereign ‘beareth the person of the people . . . beareth also his own natural person.’ When there are conflicts of interest between the artificial person representing the commonwealth and his own natural person, the sovereign prefers the private

---

2Ibid., 7.
3Ibid., 114.
4Ibid., 138.
interest to the public one, because human passions are stronger than reason.\(^1\) The reason why the subjects institute a sovereign with absolute power is to preserve their own lives and property. Can the sovereign, who cares more about his private benefit than the public one, accomplish that task? If he cannot, do the subjects have the right to rebel against the sovereign? My paper attempts to examine Hobbes’s concept of the sovereign in order to answer these questions.

According to Hobbes’s view in *Leviathan*, before they become ‘visible actions,’ all human voluntary motions, like ‘going, speaking,’ have their small interior beginnings called endeavor.\(^2\) ‘This endeavour, when it is toward something which causes it, is called APPETITE, or DESIRE . . . And when the endeavour is from onward something, it is generally called AVERSION.’\(^3\) The object of our desire is considered ‘good,’ and the object of our aversion, ‘evil.’\(^4\) When we ponder upon whether we should engage in a certain action, ‘appetites,’ ‘aversions,’ ‘hopes,’ and ‘fears’ ‘concerning one and the same thing, arise alternately,’ and all these passions will alternate and continue in our bodies until the action is completed or thought to be impossible, and this process is called by Hobbes ‘deliberation.’\(^5\) Hobbes explains that ‘it is called deliberation; because it is a putting an end to the liberty we had of doing, or omitting, according to our own appetite, or aversion,’ explains Hobbes.\(^6\) Gert observes that Hobbes’s definition ‘makes deliberation sound more like a succession of emotional states than a consideration of the consequences of the various alternative courses of actions.’\(^7\) It is the emphasis on the emotional aspect of deliberation that makes Hobbes’s concept of deliberation distinctive. Will, defined by Hobbes, ‘is the last appetite in deliberating,’ and voluntary act is what ‘proceedeth from the will.’\(^8\) Will is therefore the appetite that marks the end of our deliberation and the beginning of our voluntary actions. The object of all human voluntary acts is one’s own good.\(^9\)

For Hobbes, there is no final end for our desire. We desire one thing after another, and if we can continuously succeed ‘in obtaining those things which a man from time to time desireth,’ then that ‘continual prospering’ is ‘the felicity of this life.’\(^10\) Since ‘life itself is but motion, and can never be without desire, nor without fear,’\(^11\) ‘there is no such finis ultimus (utmost aim), nor summum bonum, (greatest good)’ for us.\(^12\) ‘Felicity is a continual progress of the desire, from one object to another; the attaining of the former, being still but the way

---

\(^{1}\)Ibid., 124.
\(^{2}\)Ibid., 34.
\(^{3}\)Ibid., 34.
\(^{4}\)Ibid., 35.
\(^{6}\)Ibid., 40.
\(^{10}\)Ibid., 41.
\(^{11}\)Ibid., 41.
\(^{12}\)Ibid., 65.
to the latter,’ observes Hobbes. In other words, we are perpetually driven by our desire for something good, and this desire will never be completely satisfied in this life. The will is therefore never free, because it has to pursue one desired object after another. The use of the word ‘free-will,’ according to Hobbes, refers to ‘the liberty of the man; which consisteth in this, that he finds no stop, in doing what he has the will, desire, or inclination to do.’

In Hobbes’s opinion, there is ‘a general inclination of all mankind, a perpetual and restless desire of power after power, that ceaseth only in death.’ The cause of our will to power arises from our uncertainty of maintaining the present ‘power and means to live well.’ We seek more power in order to keep the power we have already, and that quest for power lasts as long as we are alive, as long as we desire to live well. However, this perpetual pursuit for prosperity, this strong will for living well cannot be fulfilled unless we are willing to ‘obey a common power.’

Hobbes thinks ‘that during the time men live without a common power to keep them in awe, they are in that condition which is called war,’ which is a ‘war of every man against every man.’ In the state of nature, the human life is ‘solitary, poor, nasty, brutish, and short.’ It is not impossible for us to come out of this miserable state of war, but in order to accomplish that the natural liberty of men needs to be restrained and a common power has to be established.

Since in the state of mere nature, ‘every man has a right to every thing; even to one another’s body,’ it is evident that if ‘the absolute liberty’ in the natural state is not restrained, there is no hope for peace. Actually, the condition of absolute liberty is ‘the condition of war,’ in which neither sovereigns nor subjects exist. For Hobbes, this kind of equality undermines the foundation of peace, because he does not believe that ‘peace without subjection’ is possible. For if we could suppose a great multitude of men to consent in the observation of justice, and other laws of nature, without a common power to keep them all in awe; we might as well suppose all mankind to do the same; and then there neither would be, nor need to be any civil government, or commonwealth at all,’ states Hobbes. For him, the liberty of each individual in the state of nature, because it is absolute, brings absolute misery. Self-government is not sufficient to make people follow the moral laws, so human beings constantly have conflict with each other. Therefore, the absolute liberty of each individual does not benefit anyone.

1Ibid., 65-66.
2Ibid., 140.
3Ibid., 66.
4Ibid., 66.
5Hobbes, Leviathan, 66.
6Ibid., 84.
7Ibid., 85.
8Ibid., 84.
9Ibid., 87.
10Ibid., 235.
11Ibid., 235.
12Ibid., 112.
13Ibid., 112.
In *Philosophical Rudiments Concerning Government and Society*, Hobbes makes a crucial remark about the connection between law and liberty: ‘For the laws being removed, our liberty is absolute. This is first restrained by the natural and divine laws; the residue is bounded by the civil law.’ In *Leviathan*, Hobbes explains in details the different stages that liberty goes through to be further and further restrained. ‘LIBERTY, or FREEDOM signifieth (properly) the absence of opposition; (by opposition, I mean external impediments of motion),’ writes Hobbes.² The seed of our surrendering part of the absolute liberty has been embedded in our nature. Both passions and reason help humans to take the first step to come out of the natural condition of war and seek peace, so human beings will not all perish in the war of every man against every man. ‘The passions that incline men to peace, are fear of death; desire of such things as are necessary to commodious living; and a hope by their industry to obtain them. And reason suggesteth convenient articles of peace, upon which men may be drawn to agreement,’ states Hobbes.³ ‘These dictates of nature’ are ‘theorems’ found out by reason for our self-preservation. They are natural laws ‘if we consider the same theorems, as delivered in the word of God.’⁴ The ‘fundamental law of nature; which is, to seek peace, and follow it,’ is a dictate of reason.² From this law of nature the second law is derived: ‘that a man be willing, when others are so too, as far-forth, as for peace, and defence of himself he shall think it necessary, to lay down this right to all things; and be contented with so much liberty against other men, as he would allow other men against himself.’⁶ Hobbes explains that to ‘lay down a man’s right to any thing, is to divest himself of the liberty, of hindering another of the benefit of his own right to the same.’⁷ In other words, reason suggests that man should give up part of his right, that is, his liberty, and allow himself as much liberty against the other man as he allows the other man to have against him, that is, ‘whatsoever you require that others should do to you, that do ye to them.’⁸

The person who lays down his right simply stands out of the way of another man, so the latter ‘may enjoy his own original right, without hindrance from him.’⁹ That is to say, the person to whom the right is transferred is not impeded by the person who lays down his right. However, some rights are inalienable, for example, one’s right to ‘resist violence’ against oneself.¹⁰ ‘The mutual

---
²Hobbes, Leviathan, 139.
³Ibid., 86.
⁴Ibid., 106.
⁵Ibid., 87.
⁶Ibid., 87.
⁷Ibid., 87.
⁸Ibid., 87.
⁹Ibid., 87.
¹⁰Hobbes, Leviathan, 205.
transferring of right, is that which men call contract,’ states Hobbes.\(^1\) In the state of nature, since the common power is not erected yet, what binds the contractors to perform the contract is only the natural law. When one of the contractors performs his part of the contract and trusts that the other would perform his part ‘at some determinate time’ in the future, ‘then the contract on his part, is called PACT, or COVENANT.’\(^2\) However, the one who performs the covenant first can never be sure the other contractor will truly perform his part later, ‘because the bonds of words are too weak to bridle men’s ambition, avarice, anger, and other passions, without the fear of some coercive power.’\(^3\) Therefore, the one who performs first would ‘betray himself to his enemy; contrary to the right (he can never abandon) of defending his life, and means of living.’\(^4\) In other words, by performing the covenant first, the person puts himself in a precarious situation, which is against the right of nature, namely, ‘by all means we can, to defend ourselves.’\(^5\) This shows the necessity of establishing a common power ‘to constrain those that would otherwise violate their faith,’ so both contractors will be obliged to perform the covenants.\(^6\)

Evidently, it is essential for people to perform their covenants; otherwise, war will resume. The third law of nature is therefore crucial, that is, ‘that men perform their covenants made.’\(^7\) To break a covenant is unjust; therefore ‘the nature of justice, consisteth in keeping of valid covenants.’\(^8\) However, unless there is a common power to keep people in awe, to threaten them with punishments ‘greater than the benefit they expect by the breach of their covenant,’ people will not feel compelled to perform the covenant.\(^9\) In a commonwealth, what makes up for the right to everything that people give up is the ownership of private property. Hobbes states that the ‘coercive power . . . make[s] good that propriety, which by mutual contract men acquire, in recompense of the universal right they abandon: and such power there is none before the erection of a commonwealth.’\(^10\)

For Hobbes, human beings ‘naturally love liberty, and dominion over others.’\(^11\) In his opinion, human beings are perpetually inclined to desire after power,’ because they ‘cannot assure the power and means to live well, which he hath present, without the acquisition of more.’\(^12\) By implication, the more power humans have, the more they feel secure about their own safety and property. ‘The greatest of human powers, is that which is compounded of the powers of most men, united by consent, in on person, natural, or civil, that has

\(^{1}\)Ibid., 89.
\(^{2}\)Ibid., 89.
\(^{3}\)Ibid., 91.
\(^{4}\)Ibid., 91.
\(^{5}\)Ibid., 87.
\(^{6}\)Ibid., 91.
\(^{7}\)Ibid., 95.
\(^{8}\)Ibid., 95-96.
\(^{9}\)Ibid., 95-96.
\(^{10}\)Hobbes, Leviathan, 96.
\(^{11}\)Ibid., 111.
\(^{12}\)Ibid., 66.
the use of all their powers depending on his will; such as is the power of a commonwealth,’ observes Hobbes.\(^1\) To institute a commonwealth is to have the greatest human power, which can guarantee that we can live well. In order to erect a commonwealth, a multitude of men have to covenant with each other and to authorize one person or an assembly of men to represent them. At the same time they have to give up their right of governing themselves to this person or this assembly of men in order to be ‘united in one person,’ which ‘is called a COMMONWEALTH.’\(^2\) By doing so, the multitude of men stand ‘out of his way, that he may enjoy his own original right,’ without their ‘hindrance.’\(^3\) By means of the authority given to him ‘by every particular man in the commonwealth,’ this person can use all their ‘power and strength.’\(^4\) Thus this artificial person, who represents the multitude of men, can protect their lives and property with his unlimited power. The one who bears the artificial person is called sovereign, and everyone else is his subject.\(^5\) Because all the subjects have to obey the civil law commanded by the sovereign, their liberty is restrained.\(^6\) Hobbes compares the relationship between civil laws and the subjects to that between hedges and the travelers: ‘For the use of laws . . . is not to bind the people from all voluntary actions; but to direct and keep them in such a motion, as not to hurt themselves by their own impetuous desires, rashness or indiscretion; as hedges are set, not to stop travelers, but to keep them in their way.’\(^7\) In other words, civil laws regulate people’s voluntary actions, so they will not hurt themselves or each other.

However, in Hobbes’s opinion, the sovereign is not ‘subject to the civil laws.’\(^8\) The sovereign has the power to make and repeal laws whenever he wants; therefore, he can free himself from the subjection of the laws when they are not to his advantage to keep, so he can free himself from the constraints of the law at any time. Hobbes calls the civil laws ‘artificial chains.’\(^9\) Since the sovereign is above the civil laws and not bound by them, he becomes the freest person in the whole commonwealth. The sovereign is the person who bears the artificial person, the commonwealth. Therefore, the commonwealth is free when the subjects are bound by the civil laws. The sovereign commands and the subjects simply obey, so ‘the liberty of the commonwealth’ can be secured.\(^10\) This is how people, as subjects, can achieve the goal of living well. The Leviathan, or the sovereign with unlimited power, can have the greatest amount of strength to protect his subjects’ safety, so they can enjoy the commodious living obtained by their industry.\(^11\)

---

\(^1\)Ibid., 58.
\(^3\)Ibid., 87.
\(^4\)Ibid., 114.
\(^5\)Ibid., 114.
\(^6\)Ibid., 176.
\(^7\)Ibid., 230.
\(^9\)Ibid, 141.
\(^10\)Ibid, 142.
\(^11\)Ibid, 86.
Some scholars do not think that the sovereign is unconstrained, and they have found textual evidence to support their views. Dyzenhaus thinks that since ‘a sovereign is by definition one who governs through law,’ the supposedly absolute sovereign is limited by ‘the moral constraints of the institutions of legal order.’¹ Carmichael argues that ‘the sovereign is bound by the natural duties implied by the law of nature which are the source of the limits on the proper exercise of authority.’² Even though they have different views about how a sovereign is constrained, they both consider that the sovereign is bound by the law of nature and therefore his power is not unlimited. However, when Hobbes describes the sovereign’s power as being unlimited, he means that the sovereign’s power is not limited by the liberty of his subjects. After transferring their right to the sovereign, the subjects are ‘OBLIGED, or BOUND, not to hinder’ the sovereign.³ Some rights cannot be alienated, like the right to defend oneself against violence. Right means liberty.⁴ Even though the subject is bound not to hinder the sovereign, they have their liberty to, for instance, ‘buy, and sell . . . to choose their own abode, their own diet. . . .’⁵ After having listed the liberty of subjects, Hobbes writes:

Nevertheless we are not to understand, that by such liberty, the sovereign power of life, and death, is either abolished, or limited. For it has been already shown, that nothing the sovereign representative can do to a subject, on what pretence soever, can properly be called injustice, or injury; because every subject is author of every act the sovereign doth; so that he never wanteth right to any thing, otherwise, than as he himself is the subject of God, and bound thereby to observe the laws of nature.⁶

This is the reason why when a sovereign puts ‘an innocent subject’ to death, it is not an injury to his subject, but it is an injury to God. The subject has given the sovereign the right to do whatever pleases him, that is to say, the subject is the author of all the sovereign’s acts. Therefore, when putting an innocent subject to death, the sovereign is authorized by the innocent subject himself to do so. Since the sovereign’s authority comes from the innocent subject, this act is not an injury to the latter. Yet, the sovereign is still God’s subject and thus bound to observe the natural law, which is also the divine law. By putting an innocent subject to death, the sovereign violates the eleventh law of nature, which is equity.⁷ It is an injury to God, because, bound to observe the natural law, the sovereign nevertheless violates it.

³Ibid., 88.
⁴Hobbes, Leviathan, 86.
⁵Ibid., 141.
⁶Ibid., 141.
It might be objected that the sovereign’s power is still limited because the subject has the right to defend his own life and therefore has ‘the liberty to disobey.’ However, the subject’s liberty does not make the sovereign’s power limited because the subject is the author of all the sovereign’s acts. The subject has the right and therefore the liberty to defend himself as much as he wants, but the sovereign’s power is still not limited by the liberty of the subject. The subject’s liberty is compatible with the sovereign’s unlimited power; they do not conflict with each other.

According to Hobbes, it is to one’s advantage to live in civil society rather than in the state of nature, which is a state of war, even though the unlimited power of the sovereign seems to be oppressive. Tuck points out that, in Hobbes’s opinion, ‘if you wish to preserve yourself, then it is absurd—a logical error—to suppose that you could better preserve yourself in a situation of war than one of peace.’ However, the question is, what is the price a subject has to pay for peace? Since the sovereign is bound to observe the law of nature, he has to ‘render an account thereof to God, . . . and to none but him.’ He is not obliged or bound to his subjects to preserve their lives and property, though he is bound by the natural law to perform his duty. The problem is, if Hobbes believes that human beings will not observe the natural law ‘without a common power to keep them all in awe’ and to threaten them with ‘the terror of some punishment,’ why should the sovereign be an exception? There is no guarantee that he will perform his duty, and the subjects can only trust that the sovereign will observe the law of nature.

Sorell thinks that it is worth paying the high price for peace and safety. He makes the following comment:

Though it may seem at first that by submitting to the sovereign the many obtain their safety at a high price, it turns out that is a well-ordered state they need not suffer impoverishment, rough justice, or an intolerable curtailment of their liberty.

In the best scenario, the commonwealth might turn out to be ‘a well-ordered state,’ but what would the commonwealth look like in the worst scenario? The worst scenario, in my opinion, might be that the sovereign ignores his public duty, oppresses his subjects, and acts arbitrarily with all his unlimited power in order to satisfy his own desire. Dyzenhaus thinks that ‘a sovereign is not a natural but an artificial man,’ but that is not the case. Hobbes states that the sovereign bears both the artificial person and ‘his own natural person.’ Because human passions are stronger than reason, if ‘the common interest’ of

---

1Ibid., 144.
2Ibid., 141.
5Ibid., 112, 95.
the artificial person happens to be in conflict with ‘the private good’ of the natural person, the sovereign ‘prefers the private.’ That is to say, the sovereign will act according to his self-interest, disregard that he ought to perform his duty, and violate the natural law.

In a footnote commenting on the sovereign’s absolute power in his *Philosophical Rudiments*, Hobbes writes the following statement: ‘For he that hath strength enough to protect all, wants not sufficiency to oppress all. Here is no other difficulty then, but that human affairs cannot be without some inconvenience.’ In other words, oppression is a necessary ‘inconvenience’ that the subjects have to endure in order to be protected. Hobbes advises the readers to ‘brook with patience some inconveniences under government,’ because of war is much worse. My contention is that, in the worst case, a sovereign can use all the power and strength from his subjects to kill them and take away their property. Personal safety and the protection of private property are the two main reasons why human beings institute a sovereign. If the sovereign can take away a subject’s life or property for his self-interest without being punished by the civil law, it is counter-purposive to establish a commonwealth under his rule.

Shelton observes that since ‘the relationship between ruler and ruled is symbiotic,’ it is not to the sovereign’s advantage if he oppresses his subjects. ‘Thus, if a monarch is governed by rational self-interest he will know that for his kingdom to flourish, his subjects must also flourish,’ notes Shelton. He then asks a very important question: ‘But what if the ruler acts irrationally and, contrary to his own best interests, brings suffering to his subjects?’ I think this is the question to which Hobbes fails to provide his readers with a satisfactory answer.

In *Leviathan*, Hobbes states that ‘the name of tyranny, signifieth nothing more, nor less, than the name of sovereignty . . . saving that they that use the former word, are understood to be angry with them they call tyrants.’ Evidently, for Hobbes, all sovereigns are tyrants, but the subjects should not use such a word to express their anger toward the sovereign and be tolerated, because ‘the toleration of a professed hatred of tyranny, is a toleration of hatred to commonwealth in general,’ which can cause the dissolution of the commonwealth. The multitude of men institute a commonwealth to protect their lives and their property from being molested by other subjects, but the sovereign can take both away without being punished. For Hobbes, the sovereign is just even when he puts an innocent subject to death. ‘For . . . that nothing the sovereign representative can do to a subject . . . can properly be called injustice, or injury; because every subject is author of every act the

---

3Ibid., xxi.
5Ibid., 219.
sovereign doth,’ explains Hobbes. The sovereign is bound ‘to observe the laws of nature,’ so they only sin to God by putting an innocent subject to death.  

Hobbes also gives another reason why the sovereign can never do injury to his subjects. ‘For injury . . . is nothing else but a breach of contract; and therefore where no contracts have part, there can be no injury.’ Since the sovereign does not covenant with the subjects, he can do no injury to them.

Another inconvenience that the subjects have to endure is concerning the loss of private property. A subject ‘may be deprived of all he possesseth’ ‘for the enriching of a favourite or flatterer.’ Hobbes believes that the idea ‘that every private man has an absolute propriety in his goods; such, as excludeth the right of the sovereign’ is detrimental to a commonwealth. The sovereign needs to have the right to his subjects’ property in order to ‘perform the office they have put him into; which is, to defend them both from foreign enemies, and from the injuries of one another.’ However, it is possible for the sovereign to abuses his power at times. He can deprive his subjects of their property for his own self-interest, yet, since he does not covenant with his subjects, there is no injury done to his subjects. Actually, according to Hobbes’s definition of injury, which means injustice, as being ‘sine jure,’ if a subject refuses to let the sovereign take away his property, the subject is injurious to both the sovereign and his fellow-subjects. He is injurious to his fellow-subjects because the subjects have ‘compacted with each other to obey;’ he is injurious to the sovereign because he resumes ‘that right which he hath given him, without his consent.’

How can we trust that a sovereign, who is above the civil law and not punishable, will not abuse his power, especially when he has unlimited power and nobody has the right to stand in his way to hinder him from using his power? Punishment, according to Hobbes, is for correction. In the worst scenario, if the sovereign is above the law, he can do anything that satisfies only his desires but damages his subjects. He could abuse his subjects in many ways without being punished. Since punishment is for correction, a tyrant, who is not given a chance to correct his behavior because he is not punishable, will probably continue to oppress his subjects. Hobbes thinks that human beings in the state of nature will injure each other, so their liberty needs to be restrained by the civil laws. Why should the subjects trust that the sovereign, with endless desire for power and above the civil laws, will not oppress them?

The sovereign is only bound to the law of nature. If a person violates a natural law, according to Hobbes, a natural punishment will follow, which is

---

1Ibid., 141-42.
3Hobbes, Leviathan, 125.
4Ibid., 215
5Ibid., 215-16.
6Ibid, 88.
7Hobbes, EW, 101-02.
9Ibid., 231.
also ‘a punishment divine.’\footnote{Ibid., 206.} The natural punishment of ‘negligent government of princes’ is ‘rebellion,’ which, in turn, ‘is naturally punished . . . with slaughter.’\footnote{Ibid., 244.} Hobbes explains the concept of natural punishment as follows:

[W]hereas to certain actions, there be annexed by nature, divers hurtful consequences; as when a man is assaulting another, is himself slain, or wounded; or when he falleth into sickness by the doing of some unlawful act; such hurt, though in respect of God, who is the author of nature, it may be said to be inflicted, and therefore a punishment divine; yet it is not contained in the name of punishment in respect of men, because it is not inflicted by the authority of man.\footnote{Ibid., 206-07.}

Hobbes reminds the readers that it is God, not human beings, who is the real author of natural punishment. Some readers might still interpret the concept of natural punishment in the way not intended by Hobbes, despite his reminder. Ryan makes the following observation:

The genius of Hobbes was to produce a theory that, because it was built on individualist and rationalist foundations, must, in spite of its author’s intentions, leave room not only for individual resistance but also, \textit{in extremis}, for fully fledged revolution.\footnote{Ryan, A. (1996). ‘Hobbes’s Political Philosophy.’ T. Sorell(ed.), \textit{The Cambridge Companion to Hobbes.} Cambridge: Cambridge UP, 241.}

From Hobbes’s concept of natural punishment, it can be inferred that a tyrant will be naturally punished by rebellions if he does not govern well. Yet, Hobbes thinks that this punishment is authorized by God, not by the rebellious subjects. ‘The only time a subject can lawfully cease to obey his sovereign is when the latter is no longer able to provide security,’ states Shelton.\footnote{Shelton, \textit{Morality and Sovereignty}, 259.} Some readers might therefore reason that when the sovereign tries to put an innocent subject to death, he fails to provide security to the subjects. Therefore, the subject can exercise his inalienable right of defend himself against the violence inflicted by the sovereign. Other subjects could join him to rebel against the tyrant together because they think the sovereign fails to protect them. In other words, some readers might conclude that Hobbes actually provides a theory of revolution.

However, Shelton writes, ‘Hobbes has said there is no theoretical justification of revolution.’\footnote{Ibid., 259.} Hobbes declares his attitude towards rebellions in the following passage:

It is true, that a sovereign monarch . . . may ordain the doing of many things in pursuit of their passions, contrary to their own consciences,
which is a breach of trust, and of the law of nature; but this is not enough to authorize any subject, either to make war upon, or so much as to accuse of injustice, or any way to speak evil of their sovereign; because they have authorized all his actions, and in bestowing the sovereign power, made them their own.¹

Although Hobbes justifies tyranny rather than rebellions, it does not mean that people will not rebel against a tyrant. They might rebel--and the sovereign probably should keep that in mind--yet, the subjects do not have the right to rebel. If a rebellion actually happens, it is a divine punishment authorized by God, not by the subjects. Since the rebellion violates the natural law, it will be naturally punished by slaughter. Therefore, the question still remains: why should the subjects go through rebellions to overthrow a tyrant and then face the natural consequence of being slaughtered? The reason why people covenant with each other to establish a commonwealth is to protect their lives and property. Yet, if the sovereign has the right to take away their lives and property, to institute a Leviathan does not serve the purpose of its creation.

Hobbes is fully aware that people might imagine ‘many evil consequences that ‘so unlimited a power’ can bring, but he suggests that the subjects have to choose between the consequences of ‘perpetual war of every man against his neighbor’ and these ‘inconveniences.’² According to him, the former of these two is much worse. However, it is not necessary to choose the lesser of the two evils presented by Hobbes. My paper attempts to demonstrate that Hobbes has not sufficiently proved that instituted tyranny is the solution for the problems human beings face in the state of nature. Therefore, Hobbes’s political theory needs to be re-examined.

Bibliography


¹Hobbes, Leviathan, 165.
Towards a Kantian Anthropology: The Question of Human Being in Kant’s Groundwork and Second Critique

Alan McLuckie

In the Lectures on Logic, Kant argues that four questions collectively unite all interest of reason in its worldly or cosmological sense:

1. What can I know?
2. What ought I to do?
3. What may I hope?
4. What is man [the human being]? (Logic 538).

Although he does not explicitly offer a sustained treatment of the fourth question in his critical writings, the Lectures on Logic makes it clear that the first three relate directly to the fourth and that this last question is the driving concern behind Kant’s philosophical vision (Logic 538). The task of this paper will be to offer a reading of the Groundwork and, to a lesser extent, the second Critique with a view to highlighting both the presence of the question “what is the human being?” in these works and its significance for Kant’s philosophical venture.

In Section One I provide an overview of the philosophical division of labour insisted upon by Kant in the Preface to the Groundwork, including a brief outline of Kant’s conception of practical or pragmatic anthropology. The point I want to emphasize here is that although the Groundwork is explicitly concerned with the formal elucidation of the principle of morality entirely a priori without regard for particular human beings or human nature more generally, careful reading of the Groundwork reveals an intimate concern with the nature of humankind and its place within Kant’s ethical framework. I shall attempt to elucidate this intimacy in the Section Two, where I consider Kant’s treatment of the concept of duty in both the common and philosophical moral consciousness and its relation to the cultivation of moral character, which Kant sees as the true vocation of humankind. As will be made clearer throughout,

1The list provided in the first Critique (KrV A 805/B 833), for example, omits it entirely.
2See, e.g., G. Felicitas Munzel’s Kant’s Conception of Moral Character: The “Critical” Link of Morality, Anthropology, and Reflective Judgment.
while the formal aspect of Kant’s critical philosophy provides insight as to the formal or, so to speak, metaphysical essence of human being, there is also a subjective participation (an empirical essence, one might say) required to bring this formal element to fruition. I conclude in Section Three by offering a brief account of the moral education Kant prescribes and its role for the realization of the moral law in the world.

Division of Labour, Metaphysics and Anthropology

In the preface to the *Groundwork*, Kant insists on a division of labour for philosophy ‘so as to be able to perform it most perfectly and with greater facility’ (G 4:388). For Kant, *natural* and *ethical* philosophy each have an empirical part and a pure part: ‘All philosophy insofar as it is based on grounds of experience can be called empirical; but insofar as it sets forth teachings simply from a priori principles it can be called pure philosophy’ (G 4:388). The division proper to ethics, Kant argues, is as follows: ‘the empirical part might be given the special name practical anthropology, while the rational part might properly be called morals’ (G 4:388). Failure to adhere to this division in labour, Kant proclaims, does the greatest disservice to philosophy. As Kant puts it, ‘Where work is not so differentiated and divided, where everyone is a jack-of-all-trades, there trades remain in the greatest barbarism’ (G 4:388).

Kant tells us that ‘The present groundwork is … nothing more than the search for and establishment of the supreme principle of morality, which constitutes by itself a business that in its purpose is complete and to be kept apart from every other moral investigation’ (G 4:392). The *Groundwork* thus explicitly represents the formal part of ethical philosophy, responsible for elucidating the formal principle of morality entirely a priori. To make his case, Kant contends that ‘Everyone must grant that a law, if it is to hold morally, that is, as a ground of obligation, must carry with it absolute necessity… therefore, the ground of obligation here must not be sought in the nature of the human being or in the circumstances of the world in which he is placed, but a priori simply in concepts of pure reason’ (G 4:389, my emphasis). Kant’s emphasis on the formal aspect of ethical philosophy might justly lead one to be suspicious of Kant’s ethics to the extent that it seems to exclude elements of humanity we might intuitively think essential to ethics. Dudley, for example, accuses Kant of failing to account for the various social and political situations within which humanity finds itself (Dudley 5-7). However, Kant’s insistence on the philosophical division of labour in the *Groundwork* suggests that there is, after all, “something other” than the search for and establishment of the supreme principle of morality at play here.

The “nothing more” in the passage just quoted in the last paragraph above is spurious. That is, although the (explicit) primary concern is with the pure aspect of morality, the *Groundwork* is not exclusively concerned with this formal aspect of the moral law. On the contrary, there is a sense in which (at least implicitly) the primary concern in the *Groundwork* really is the empirical
Towards a Kantian Anthropology: The Question of Human Being in Kant’s Groundwork and Second Critique

aspect: namely, the finite rational being who wills the moral law.¹ When the principle of morality ‘is applied to the human being,’ *qua* finite rational agent, Kant argues, ‘it does not borrow the least thing from acquaintance with him (from anthropology) but gives to him, as a rational being, laws a priori, which no doubt still require a judgment sharpened by experience, partly to distinguish in what cases they are applicable and partly to provide them with access to the will of the human being and efficacy for his fulfillment of them’ (G 4:389). There are two issues here: the judgment sharpened by experience to apply the law, and the need over time to shape one’s inclinations and habits to morality. As the ground of obligation, the moral law is binding for all human beings independent of experience with or knowledge of them and their particular, accidental natures. Nevertheless, when the principle of morality is applied to human beings in *actual* cases, acquaintance with human beings in their particularities is required. What is needed to apply the moral law is the empirical part of ethical philosophy, which, for Kant, is practical anthropology.²

It seems that all sorts of empirical knowledge are required for the application of moral principles, including laws of nature, and the contingencies of the specific circumstances in which humanity finds itself, etc. Kant’s conception of a practical anthropology, however, is importantly a shift from a *mere* empirical study of the human being or human nature to the *inclusion* of teleological considerations. Kant makes a distinction between physiological and practical anthropology: the former studies human beings historically (what humankind *has been* in all of its particularities) versus what human being *qua* free agent *can and ought to become*. On Kant’s view, what is common and permanent in human nature is not to be found in what humankind or the individual human being *has been*, but rather in what it *can make itself to be*.³ Much more, of course, needs to be said about Kant’s conception of anthropology. The point I want to emphasize for our purposes here, however, is that although the *Groundwork* is explicitly concerned with the *formal*

¹It is also important to note that the “search for” the moral law is not meant to imply the search for some truth “out there” awaiting our discovery, nor is the “establishment” meant to suggest that Kant intends to “make it up.” Rather, as we shall see in Section Two below, it is the search for and establishment in the sense of making explicit what we always already know in moral experience.
²Wood notes four distinct, though related, senses in which to understand Kant’s “pragmatic” anthropology: pragmatic versus physiological; pragmatic versus scholastic; pragmatic as useful; pragmatic as prudential (Wood 2003, 40-42). Due to restrictions of both space and time, I shall only be treating the first sense of pragmatic anthropology in this paper.
³Van De Pitte, for example, observes that empirical anthropology ‘can, of course, only describe man as he has expressed his nature in the course of historical events. But this will only give us the various roles that man has played, the various masks that he has worn from time to time. We will still know nothing of his essential nature, and of what man ought to be. Kant’s formulation, on the other hand, is a prescriptive, and even a *creative* anthropology – it emphasises man’s responsibility to become what he can be, that is, to fulfill his potential (Van De Pitte xxi).
elucidation of the principle of morality entirely \textit{a priori} without regard for particular human beings or human nature more generally, there is also a subjective participation required to bring this formal element to fruition, which suggests concern for the embodied and socially-embedded agent.

\section*{The Principle of Duty and the Cultivation of Moral Character}

Section I of the \textit{Groundwork} proceeds analytically from a common moral consciousness to a sophisticated philosophical moral consciousness of the principle of duty. Kant’s procedure is Socratic in the sense that, ‘without in the least teaching it anything new, we only, as did Socrates, make [duty] attentive to its own principle (G 4:404, my emphasis). On Kant’s view, common moral cognition, as a matter of course, makes the distinction between inclination and obligation. As agents forced to act in the world we do not need philosophy to tell us that we are morally obligated: we \textit{always already know} we are morally obligated precisely because we cannot but make this distinction in our everyday interactions within the world. The primary moral question for Kant is thus not “Why be moral?” but rather “What ought I to do?” Philosophy does not teach the common moral consciousness anything new, but it does ‘provide access and durability for its precepts’ (G 4:05). The task of philosophy, for Kant, is here nothing more than the \textit{explication} and \textit{legitimization} of a knowledge we \textit{always already} have in moral experience itself. On Kant’s view, philosophy is so required because, as human beings, we are inescapably subject to inclinations that present a ‘powerful counterweight to all the commands of duty’ (G 4:405). These inclinations are simply \textit{given} in nature are thus intrinsic to human being. As such, these “innocent” inclinations cannot be simply “given up” or “transcended” (KpV 5:84). The problem, however, is that the inclinations accidental to one’s particular being often influence the choice of maxims upon which one acts and can thus override one’s sense of moral obligation.

Kant refers to this tension between inclination and obligation as the \textit{natural dialectic}: a propensity to rationalize against those strict laws of duty and to cast doubt upon their validity, or at least upon their purity and strictness, and, where possible, to make them better suited to our wishes and inclinations, that is, to corrupt them at their basis and to destroy all their dignity – something that even common practical reason cannot, in the end, call good (G 4:405). It is in this sense that philosophy is needed for the legitimization of morality, since both “the powerful counterweight of our inclinations” as well as the natural dialectic to which they give rise undermine and corrupt the very ground of morality. Allowing oneself to be guided by the inclinations represents, for Kant, a perversion of reason to the extent that, rather than attending to the principle of duty and law giving, reason here is used ‘only to look after the interests of the inclinations, whether singly or, at most, in their greatest compatibility with one another’ (G 4:407). It is through a transition from the common moral consciousness to a sophisticated philosophical moral consciousness that
philosophy, as Kant understands it, undertakes the legitimization of the principle of duty and guards against this natural dialectic into which reason would otherwise inevitably fall.

Both the common and philosophical moral consciousness, Kant insists, agree on the following representation of the principle of duty: ‘*I ought never to act except in such a way that I could also will that my maxim should become a universal law*’ (G 4:402, Kant’s italics). It is interesting to note that Kant thinks that even moral skeptics do ‘not … call into doubt the correctness of the concept of morality but rather [speak] with deep regret of the frailty and impurity of human nature, which is indeed noble enough to take as its precept an idea so worthy of respect but at the same time is too weak to follow it’ (G 4:406). Even those who are skeptical about the ability of humanity to live up to the principle of duty do not in the end deny the veracity of (and therefore the obligation to) the moral principle itself. It is from this principle of duty that Kant derives the moral law. Although treatment of the deduction of the moral law itself will remain beyond the scope of this paper, it will be useful to sketch out the three formulations and two sub-formulations of the principle of morality outlined by Kant in section II of the *Groundwork*.¹

The first formulation of the principle of morality, the Formula of Universal Law, is effectively the principle of duty noted above: ‘*act only in accordance with that maxim through which you can at the same time will that it become a universal law*’ (G 4:421, Kant’s italics). The Formula of the Law of Nature is a variant of the Formula of Universal Law: ‘*act as if the maxim of your action were to become by your will a universal law of nature*’ (G 4:421, Kant’s italics). The second formulation, the Formula of Humanity as an End in Itself, says: ‘*So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means*’ (G 4:429, Kant’s italics). The third formulation, the Formula of Autonomy, says: ‘choose only in such a way that the maxims of your choice are also included as universal law in the same volition’ (G 4:440). The Formula of the Realm of Ends is a variant of the Formula of Autonomy: ‘act in accordance with the maxims of a member giving universal laws for a merely possible kingdom of ends’ (G 4:439). Crucial to notice here is that the various representations of this principle ‘are at bottom only so many formulae of the very same law, and any one of them unites the other two in it’ (G 4:436). Nevertheless, the Formula of Autonomy, we shall see presently, is the penultimate formula of the principle of morality for Kant and represents an ontological claim regarding human being: to be human is to be a moral agent.

¹While there is debate in the literature about how many formulations and variations of the principle of morality Kant actually presents, I am here concerned with what they collectively say about human being for Kant. For the purposes of summary exegesis I have followed Allen Wood’s succinct presentation of the various formulations as outlined in his book *Kant’s Ethical Thought* (Wood 2003, 17-8).
²The “ultimate” formulation is how one effectively realizes the moral law in the world. I will have more to say about this issue below.
For Kant, human beings—and indeed all finite rational beings—can be considered from two distinct standpoints: as belonging to the intelligible world (the pure, rational world of understanding) and as belonging to the sensible world (the empirical world). These two standpoints capture two distinct natures that are nonetheless united in one being. In effect, these two viewpoints reflect the philosophical division of labour discussed above. That is, the human being has a pure, rational element as well as an empirically conditioned element. Following through with this analogy, morality must carry with it absolute necessity and thus must be cleansed of anything empirical. The Formula of Universal Law does just this, providing the form of universality that is without empirical content.

To be, for Kant, is to exist as determined under rules. There are rules of nature and rules of freedom, and so two senses of being for Kant. Kant argues in *Groundwork* III that the ‘Will is a kind of causality of living beings insofar as they are rational, and freedom would be that property of such causality that it can be efficient independently of alien causes determining it, just as natural necessity is the property of the causality of all nonrational beings to be determined to activity by the influence of alien causes’ (G 4:446). Thus, ‘a free will and a will under moral laws are one and the same’ (G 4:447). Considered from the standpoint of sensibility alone, human beings are determined by natural necessity (the a priori causal laws in nature) in the same way as all non-rational beings are determined. It is the standpoint of the intelligible world, Kant claims, that provides insight into the ability of finite rational agents to act as a causal force independent of the influences of sensibility. This, for Kant, is freedom in the negative sense and, significantly, it points to a positive conception of freedom, namely the will’s imposing its own causality onto nature. On Kant’s view, positive freedom just is to make the Formula of Universal Law effectively the Formula of a Law of Nature. That is, since nature is unequivocally subject to causal laws, the positive conception of freedom imparts laws upon nature to which she is then inescapably bound. Notice, importantly, that the Formula of the Law of Nature itself contains no empirical import and thus remains formal.

For Kant, all action contains and acts upon an end. That is, to act is to do so for some purpose. In this respect the Formula of Humanity serves ‘as the limiting condition of all merely relative and arbitrary ends’ (G 4:436). In other words, the Formula of Humanity ensures the absolute equal worth of all human beings, since it remains abstracted from any sensible influence and is not concerned with this or that human being but with humankind as a whole. As

---

1 Though a full treatment of the distinction is beyond the scope of this paper, it is important to note that the “will” referred to here is *Wille* as opposed to *Willkür*. Although, there is sometimes slippage in Kant’s usage of the distinction, *Wille* is the pure will as capacity to give the moral law to oneself (independent of any inclinations of sensibility) and *Willkür* represents the capacity of choice in choosing between competing maxims. The distinction is important in that if all free choices were one and the same with the moral law (that is, if all free choices were merely *Willkür*), then one could not be held morally culpable, since the free choice would not be absolutely independent of heteronymous ends and one could not be said to be obligated in the sense required for morality: All free choices would be indifferently the same.
we shall see, however, what is key for Kant is how the Formula of Autonomy unites both (the sensible and intelligible) aspects of human nature.

The autonomous agent, for Kant, just is the finite rational being with a free will subject to the moral law. As Kant puts it in the second Critique, the law of autonomy ‘is the moral law, which is therefore the fundamental law of a supersensible nature and of a pure world of the understanding, the counterpart of which is to exist in the sensible world but without infringing upon its laws’ (KpV 5:43). Key to notice here is that Kant says “which is to exist” and not “exists.” This is instructive because it emphasizes the role of the human being in relation to the effective reality of the moral law: All human beings are themselves bound by, and so responsible for, the project of bringing the moral law into the sensible world. It is crucial to note that although human beings are necessitated by the moral law since they are fundamentally autonomous beings, the effective reality of the moral law just is what is willed into the sensible world by human beings. In other words, how one approximates the moral law is the effective reality (and thus objective, though only practical, reality) of the moral law. After all, approximating the moral law is all a sensibly affected being can reasonably hope to achieve, since one cannot simply give up or transcend one’s inclinations outright. As Kant puts it in the second Critique, ‘being a creature and thus always dependent with regard to what he requires for complete satisfaction with his condition, [a human being] can never be altogether free from desires and inclinations’ (KpV 5:84). For Kant, happiness is the complete satisfaction with one’s condition and virtue is worthiness to be happy. Moral being is brought to fruition in the highest good, which is the commensuration of virtue and happiness in autonomous agents. Moral being, then, requires that we, so to speak, assimilate our inclinations, rather than simply giving them up. In other words, our inclinations are to be brought into conformity with and not eliminated from moral being. Human beings, as finite rational beings necessitated by the form of the moral law, give content to the moral law in and through moral action. This content, in turn, constitutes the effective reality of the moral law. The Formula of Autonomy, for Kant, is what unites the form of the principle of morality with nature, since it furnishes ‘the sensible world, as a sensible nature ..., with the form of a world of the understanding, that is, of a supersensible nature, though without infringing upon the mechanism of the former’ (KpV 5:43). Put differently, the Formula of Autonomy brings the moral law into the world by changing ‘its transcendent use into an immanent use’ (KpV 5:48). As alluded to above, however, realizing the moral law in the world requires ‘moral cultivation and exercise’ (KpV 5:161), and is, on Kant’s view, the true vocation of humankind.

Towards a Kantian Moral Education

Though, regrettably, an in-depth treatment of the moral education Kant prescribes is beyond the scope of this paper, a few words might be helpful to
shed light on Kant’s procedure. For Kant, moral education consists in the cultivation of moral character. This cultivation of character does not aim to teach anything new, but rather strives to make one attentive to the principles of morality intrinsic to one’s being. Put differently, moral education strives to make one attentive to and become what one always already is. This cultivation aims to make the supreme moral principle immanent in the world and it is through the development of moral character that Kant suggests human freedom\(^1\) is realized in this world. As Munzel comments: ‘Character (or more precisely, a singular character) is the notion in terms of which Kant himself poses the question of the unity of forms of thought and life, that character is the primary instance of the way in which freedom and nature are related for him’ (Munzel 3). Furthermore, the realization of character, for Kant, ‘the formative activity peculiar to human, rational beings in relation to living nature (ApH 321). It is an activity that concretely actualizes moral law in the world, imparting its form to sensibility and effecting literally a counterimage of the objective law under the conditions of the latter. In realizing this counterimage, character makes its appearance in the world’ (Munzel 9). This appearance of character in the world, for Kant, is the realization of human being. That is, the realization of moral character, the task proper to all finite rational beings in general and human beings in particular, just is the fulfillment of what it means to be human.

The moral education Kant presents us with is one which both elucidates the formal principle of morality and attempts to guide us by providing examples of pure moral worth. These examples of pure moral worth ‘serve only for encouragement, that is, they put beyond doubt the practicability of what the law commands and make intuitive what the practical rule expresses more generally, but they can never justify setting aside their true original, which lies in reason, and guiding oneself by examples’ (G 4:409, my emphasis). That is, these examples serve to show that acting from pure duty, and not simply in conformity with duty, is something that can be and is made immanent in this world. It is something that we, as autonomous agents in the world bound by and obligated to the moral law, can hope to achieve in concreto. The principle of morality is thus not some chimerical ideal that is, practically speaking, impossible to realize for autonomous agents. Rather, these representations of pure virtue have tremendous power over the mind ‘and can provide a far stronger incentive to effect even that legality of actions and to bring forth stronger resolutions to prefer the law to every other consideration, from pure respect for it, than all the deceptive allurements of enjoyment and, in general, everything that may be counted as happiness, or even all the threats of pain and troubles can produce’ (KpV 5:152).\(^2\) Examples of pure virtue, Kant contends,

---

\(^1\)Recalling from above that “a free will and a will under moral laws are one and the same” (G 4:447) and that freedom is ‘the keystone of the whole structure of a system of pure reason’ (KpV 5:3).

\(^2\)Here, again, we should acknowledge Kant’s debt to Socrates: ‘Then when we were discussing the nature of absolute justice and injustice – as well as the perfectly just man (supposing that he exists) and the completely unjust man – we were looking for ideals and patterns of instruction.
also inevitably lead to an elevation of our soul: ‘before a humble common man in whom I perceive uprightness of character in a higher degree than I am aware of in myself my spirit bows, whether I want it or whether I hold my head ever so high, that he may not overlook my superior position. Why is this? His example holds before me a law that strikes down my self-conceit when I compare it with my conduct, and I see observance of that law and hence its practicability proved before me in fact’ (KpV 5:77). Seeing the practicability of the moral law through these concrete examples, we are ourselves moved to appropriate them into our own being and cultivate our own moral character. In so doing, we thereby bring ourselves closer to approximating the holiness of moral being in the world, giving form to freedom in this world and respecting the dignity of each and every human being.

Conclusion

The Formula of Autonomy is, for Kant, an ontological claim about human nature: to be human is to be a moral agent. That is, human beings are fundamentally autonomous agents, subject to and necessitated by the commands of morality. The definitive task of our vocation as members of humanity is to cultivate moral character so as to give objective reality, though only in a practical respect, to the moral law. As human beings we are, as a matter of course, sensibly affected. As we have seen, however, our inclinations cannot and ought not to be given up. The principle of morality, ‘That law of all laws, therefore, like all the moral precepts of the Gospel, presents the moral disposition in its complete perfection, in such a way that as an ideal of holiness it is not attainable by any creature but is yet the archetype which we should strive to approach and resemble in an uninterrupted but endless progress’ (KpV 5:83). For all finite rational beings, the archetype of the principle of morality can only be approximated and how we exact the moral law in the world just is its effective reality for us (in a practical respect). Though we can only approximate the moral law, we are nevertheless obligated to strive to approach and resemble the moral law in its archetypal form.

Because Kant’s philosophical project, on the reading I have been suggesting, represents a positive finite philosophy that is chiefly concerned with meaningful engagement within this world, it is thus also first and foremost concerned with the question “What is the human being?” Human freedom, I contend, represents the crux of Kant’s philosophy. As autonomous agents, all

We wanted to bring them into focus as models so that we might judge our own happiness or unhappiness according to the standards they set and according to the degree we reflect on them. It was not our purpose to demonstrate the possibility of fully realizing these ideals’ (Republic 472c-d, my emphasis).

Nietzsche puts this same point nicely when asking why the most powerful humans have bowed before the saint: ‘they sensed a new power, a strange, as yet unconquered enemy—it was the “will to power” that made them stop before the saint. They had to ask him—’ (BGE 51).
human beings are of absolute equal worth. As we have seen, however, the accidental natures of human beings in their particularities often obscure and thereby infringe upon this equality. Practical philosophy, for Kant, thus seeks to make what is transcendent, (the equality of all finite rational beings) immanent (the realization of human freedom in this world). Kant’s philosophy is above all one about meaning and is chiefly concerned with the true vocation of human kind: the realization of human freedom. Human freedom is only possible through our moral action in the world. It is in and through moral being that we not only give form to freedom in this world, but through the cultivation of moral character we ourselves give meaning and dignity the each and every human being. The task of philosophy, for Kant, is thus no longer to escape the cave within which we are imprisoned in favour of some transcendent ideal; rather, it is to give meaning and purpose to this life and this world, to realize that what truly imprisons us is the desire to escape the cave at all.¹

Work Cited


¹I would like to express my gratitude to Dr. Robert Burch for his inspiration and guidance through this project and to Lindsey Affleck and Ben Miller for their invaluable editorial comments. I would also like to thank the participants at the Stanford Graduate Students’ Workshop, especially Assaf Sharon, for their helpful comments and criticisms on this paper.
In recent years, the conflict between analytic and continental philosophy has taken a compelling turn. These very different traditions are finding common ground in Hegel. I say “compelling” because, although continental thinkers have long recognized Hegel’s influence in the fields of critical theory, political philosophy, and the philosophy of history, analytic thinkers have either paid little attention to Hegel or looked at him with a kind of distant disdain. But there are now signs that this is starting to change; and that even a bridge between these sometimes antagonistic traditions is being built with Hegel’s dialectic. For example, a movement toward the continental tradition from the analytic side can now be seen in the work of Robert Brandom at the University of Pittsburgh. In his essay, ‘Some Pragmatist Themes in Hegel’s Idealism,’ Brandom explores whether his own semantic pragmatist thesis that the content of empirical concepts is determined by their use ‘might best be understood by modeling them on [Hegel’s idealist thesis]’ that “the structure and unity of the concept is the same as the structure and unity of the self.” Conversely, a movement toward the analytic tradition from the continental side can be found in the work of Pirmin Stekeler-Weithofer at Universität Leipzig. In his book, Hegels analytische Philosophie: Die Wissenschaft der Logik als kritische Theorie der Bedeutung, Stekeler-Weithofer argues that Hegel’s dialectic elucidates the fundamental forms and structures of linguistic meaning through a critique of their dogmatic acceptance by what Hegel calls the ‘Understanding.’ It is my contention that the coming together of these traditions is not coincidental, but has its roots in the logic of the dialectic; specifically, Hegel supplies a common ground for these disparate traditions by preserving the analytic within the dialectic.

A recent and significant book that deals with this topic is Tom Rockmore’s Hegel, Idealism, and Analytic Philosophy. The primary thesis of this work is that the analytic and continental traditions are incompatible. The implication is that any attempt to bridge them with Hegel’s dialectic would be as much misguided as futile. Rockmore traces the incompatibility back to Kant. He argues that these traditions responded in fundamentally different ways to transcendental idealism and, as a consequence, cannot be reduced to ‘a single

1Brandom, p. 164, my brackets
common denominator\(^1\) or shared philosophical commitment. The analytic side distanced itself from Kant’s idealism and, for the most part, turned toward realism; whereas the continental side tended to accept it by developing what Rockmore describes as the “constructivism” of the Copernican revolution. There can be no bridge between the two traditions, according to Rockmore, because there is no way to reconcile realism and idealism. As he frames it, the realist assumption that to know an object is to “uncover, discover or reveal”\(^2\) a mind-independent world is not compatible with the idealist assumption that to know an object is to construct the conceptual framework for cognizing objects as they occur within consciousness (or the mind) itself.\(^3\) Of course, Rockmore’s takes a side: his aim is to defend a uniquely continental approach to philosophy, in particular, to epistemology, based on a constructionist model not at all adverse to the idealist tradition renounced by analytic philosophy. He concludes his work by pointing out the relevance of Hegel’s brand of constructivism to the contemporary problems of knowledge. He does not advise a full scale return to Hegel, as if he could cure all the ills of present-day epistemology, but believes that his constructivism carries the important insight that the theoretical knowledge of objects contains a practical element. In his view, Hegel improved upon Kant’s constructivism by abandoning the rigid distinction between theory and practice in favor of a dialectical approach in which the knowledge of objects is the result of experience, that is, of praxis. Ultimately, Rockmore believes that Hegel’s dialectical constructivism, which cannot be grafted onto the philosophical convictions of the analytic tradition, provides the preliminary conceptual framework for a continental epistemology concerned with ‘how ordinary people go about their lives’.\(^4\)

Although Rockmore presents a strong case for the incompatibility of the analytic and continental traditions, his assertion that they cannot be reduced to ‘a single common denominator’ may nevertheless be challenged. If, contrary to Rockmore, a common denominator can indeed be found, then the related charge that Hegel’s dialectic is unable to bridge the two traditions may also be disputed. In fact, the following will show that there is a common denominator shared by the two traditions, which is not readily evident in the debate between realism and idealism,\(^5\) but which can be found in the context of logic. What is more, it is argued that this common denominator is a central component of Hegel’s dialectic, giving it the unique ability to reconcile the analytic and continental traditions. This may initially seem to be the least likely place to find any commonality. From the analytic point of view, probably the most troubling aspect of Hegel’s entire philosophy is his logic because of the

---

\(^1\) Rockmore, p. 224
\(^2\) Ibid, p. 211
\(^3\) Ibid, p. 225-28
\(^4\) Rockmore, p. 228
\(^5\) For an interpretation of Hegel’s idealism that counters Rockmore’s incompatibility thesis in the context of realism and idealism, see my article ‘Reconciliation in Hegel’s Idealism’ (forthcoming publication in the Villanova journal of philosophy, Epoche, fall 2008). Here, it is argued that Hegel’s dialectical idealism attempts to reconcile the opposition between realism and idealism, which is examined through the specific examples of Aristotle and Kant.
prominent role of contradiction. Although, on the surface, it would not be at all controversial to characterize Hegel’s dialectic as a logic of motion, it becomes so once the relationship between motion and contradiction is laid out on the table. Hegel not only asserts that motion entails a contradiction, which is troubling enough, but he also champions the stronger claim that the source of all motion, i.e. that which both explains what it is and functions as its determining ground, is contradiction itself:

Contradiction is the source of all movement and vitality.¹

Something moves, not because at one moment it is here and at another there, but because at one and the same moment it is here and not here, because in this ‘here,’ it at once is and is not …

Motion is existent contradiction itself.²

Contradiction pervades every aspect of Hegel’s dialectic. This is problematic for a host of reasons, but perhaps the most significant is that, as far back as the Ancient Greeks and, most explicitly Aristotle,³ contradiction has been viewed as meaningless, incoherent and, in a word, false. The inverse significance of this strictly negative value of contradiction is that a necessary condition of all meaning, coherency, and truth is non-contradiction. This necessity is expressed formally in the Law of Non-Contradiction (LNC). In general, the LNC unequivocally denies the truth of contradiction; in particular, it asserts that no pair of contradictory opposites can be both true, ~ (α · ~α).

Thus, Hegel’s dialectic seems to be the least likely place to find a common denominator shared by the two traditions for the reason that the analytic side does not accept contradiction as a legitimate principle of logic. Although quite a few continental thinkers such as Marx, Kierkegaard, members of the Frankfurter Schule, Derrida, and others have been positively influenced by Hegel’s concept of contradiction, virtually no analytic thinker has.

From the other side, continental thinkers sympathetic to Hegel would also find the idea that the dialectic contains a common denominator to reconcile them with their analytic counterparts difficult to imagine. Not only does Hegel often refer to the analytic with some degree of derision; but it would also be fair to say that, for him, the distinction between non-contradiction and contradiction represents the conflict (not the concord) between the analytic and dialectic. Even Hegel himself seems to deny that analytic logic shares anything in common with his dialectic. In the broadest possible terms, what Hegel means by “analytic” is a logical system that denies contradiction in accordance with the LNC. By “dialectic”, in contrast, he means a logical system that affirms contradiction. Hegel develops this distinction in many ways, but one of the more conspicuous and often discussed is the difference between the Understanding (Verstand) and Reason (Vernunft). For Hegel, the

¹Hegel, Science of Logic, p. 439
²Ibid., p. 440
Understanding and Reason are not only cognitive faculties of human subjectivity, but also logical standpoints that correspond to the analytic and the dialectic, i.e. to the denial and the affirmation of contradiction respectively. The manner in which Hegel deals with the Understanding and Reason is a complex issue that has spawned a variety of plausible interpretations. Although there is almost unanimous consent that he favors Reason over the Understanding, there are a number of ways to construe this. One could argue, for instance, that Hegel’s preference for Reason represents the attempt to overcome what he considers to be the inadequacies of the Understanding and its denial of contradiction. In this case, Hegel’s dialectic could be interpreted as an attempt to abandon the analytic in order to replace it with a logic that affirms contradiction. According to this view, there is no place for the analytic in Hegel’s dialectic, which would preclude the possibility of finding a common denominator therein that could bridge the analytic and continental traditions.

Another possible reading, which the following defends, is that the Understanding is not abandoned by, but preserved within the faculty of Reason. The idea of preservation introduces Hegel’s well-known concept of aufheben, according to which anything shown to be inadequate by means of critique is not entirely disposed of, but incorporated into a more comprehensive standpoint. The preserving function of aufheben is supposed to operate in all contexts, including logic. The suggestion is that Hegel’s dialectic similarly does not abandon the analytic, but preserves it. Regardless of whether one agrees with Hegel or not, it stands to reason that he did, indeed, see a place for the analytic in the dialectic. If, in fact, one looks to what Rockmore locates as the common origin of the analytic and continental traditions – namely to Kant – then one discovers that the analytic plays an indispensable role in Hegel’s dialectic. But to see this, one must turn to a different part of Kant’s transcendental philosophy: not to his positive theory of idealism, as Rockmore does, but to his critique of Reason in the antinomies. An examination of the influence that Kant’s antinomies had on Hegel shows that the indispensability of the analytic extends to the LNC. Ironically, the logic of the antinomies influenced Hegel so extensively that the LNC became one of the central components of his own dialectic. To my knowledge, Stekeler-Weithofer is one of the only scholars who recognizes, albeit for other purposes, Hegel’s non-derisive, positive appropriation of the LNC: ‘This is the method of every “dialectical” analysis of concepts. Its goal is the restitution of the validity of the principle of contradiction’.\footnote{Stekeler-Weithofer, ‘Verstand und Vernunft’, p. 167, my translation} The common denominator overlooked by Rockmore is the analytic LNC. It is this law that can function as the cornerstone of the bridge between the analytic and continental traditions. One might immediately and, to some extent, rightly point out the triviality of this claim. The extension of the LNC is so broad that just about anything could ultimately be reconciled with just about anything else, so long as both adhere to the criterion of non-contradiction. Why stop with the analytic and continental traditions in philosophy? Why not also bridge the arbitrary gap between your
local telephone book and the menu of your favorite restaurant, so long as both adhere to the LNC? This suspicion would be entirely justified, if Hegel’s interest in the LNC was to use it as a formal standard of comparison, i.e. as the kind of standard he critically describes as “indifferent.” But this is not the case. What interests Hegel is the distinctive role of the LNC in the logic of the antinomies with specific regard to contradictory opposites. To be sure, the broad reach of the LNC renders it trivially true; but, for Hegel, the very same analytic law procures a unique significance, as Plato long ago illustrated in the Parmenides, when applied in the two-sided manner of an antinomy.

**Hegel’s Dialectic as a Logic of Antinomies**

One of the main influences on Hegel’s dialectic is Kant’s account of the antinomies in the first Critique. Kant’s influence was not his unfavorable assessment of the antinomies or his critical solution. Hegel had serious reservations about both and soundly criticized them. Irrespective of Kant’s own purposes, Hegel’s dialectic was shaped, in large part, by the logic of the antinomies. Hegel makes this quite clear in Part I of the Encyclopedia of the Philosophical Sciences, when he writes:

> It may also be remarked that, as a result of his failure to study the antinomy in more depth, Kant brings forward only four antinomies...The main point that has to be made is that antinomy is found not only in the four particular objects taken from cosmology, but rather in all objects of all kinds, in all representations, concepts, and ideas. To know this, and to be cognizant of this property of objects, belongs to what is essential in philosophical study; this is the property that constitutes what will determine itself in due course as the dialectical.¹

Kant’s failure was not that he characterized the dialectic as antinomial, but that he did not fully grasp what he had discovered. In Hegel’s view, Kant had shown that the conclusion of an antinomy is contradictory, and thus dialectical, for the reason that the thesis and the antithesis can both be justified in accordance with analytic arguments. Each side of an antinomy employs a *reductio ad absurdum* argument in which the aim is indirectly to prove the truth of something by demonstrating that its denial entails a contradiction. To be sure, the law of excluded middle confirms the conclusion of a *reductio* argument on the grounds that the falsity of the denial logically necessitates that the opposing affirmation must be true. However, the LNC plays a central role in the successful execution of the *reductio*: this kind of argument is successful if and only if the denial of what is to be indirectly proven violates the LNC. In the context of the antinomies, Kant understands the LNC as a formal law that regulates the correctness of the *reductio* irrespective of content. The LNC has

¹Hegel, *Encyclopedia Logic*, §48, Hegel’s italics
the same significance and function regardless of whether the content is the quantitative beginning in time and a limit to space, the divisibility of a thing into simple parts, the qualitative constitution of causality, or the necessity of the Absolute Being. Thus, for Kant, the LNC is a necessary condition of only the formal validity of the *reductio*: as a logical argument, the *reductio* requires that the denial of an hypothesis, claim, etc., (whatever the content) violate the LNC in order to prove indirectly that the opposing affirmation is true.

One of the consequences of the dual application of the *reductio* is that both the thesis and the antithesis, independently of one another, are each analytic arguments that abide by the LNC. Each side of the antinomy indirectly demonstrates that its own claim adheres to the standard of non-contradiction over against the failure of the other side to do so. Thus, in the case of the third antinomy, for example, the claim of the thesis that there is free causality is just as non-contradictory as the claim of antithesis that there is no free, but only natural causality. What Kant sees as the problem with the antinomies (and not just the third), Hegel sees as their merit: when both of the analytic, non-contradictory arguments are taken together as a whole, the result is contradiction, e.g. there is and is not free causality. This outcome inspired Kant to give an extensive transcendental critique of Reason that seeks to remedy the antinomies. Hegel had a different reaction. What he finds compelling is the logical peculiarity of the antinomies in which contradictory opposites adhere equally to one and the same LNC. For Hegel, the genuine import of the antinomies is what they show about this analytic law; specifically, the LNC is able to demonstrate contradiction when applied in the dual form of an antinomy. Ironically, this is not done by rejecting, but by adhering to the LNC: the thesis has a claim to legitimacy on the grounds that it is non-contradictory; conversely, the antithesis has an equal claim to legitimacy for the same reason. Due to this, Hegel came to realize that the LNC has a dialectical significance insofar as it can demonstrate contradiction; what is more, since this kind of demonstration is not accomplished by violating, but by upholding the LNC, the dialectic preserves its analytic validity. Therefore, one of the central insights that Hegel takes from Kant’s antinomies is that the LNC itself becomes a dialectical principle when applied in the two-sided manner of an antinomy, but *without losing its analyticity*.

Of course, to claim that Hegel was positively influenced by the Kantian antinomies is not to suggest that he fully agrees with him. On the contrary, Hegel disagrees with many aspects of his account, especially, for our purposes, his characterizations of the *reductio* and the LNC. Hegel criticizes both as instances of Kant’s *formalism*, which is a complex issue containing more than one defining characteristic. In the present context, the two most important are (1) that neither the LNC nor the *reductio* are constitutive of the content of the argument; and (2) both are fixed logical determinations with no trace of movement. In line with his concept of *aufheben*, Hegel does not reject the LNC or the *reductio*, but preserves them as integral components of the dialectic by exposing what he considers to be their authentic dynamic nature. For Hegel, the *reductio* is not simply a ready-made formal argument that can be
indifferently applied irrespective of content, but is a movement of the content itself, specifically, a transition of opposites, facilitated by the LNC. When, in Kant’s third antinomy, freedom is shown to contradict natural causality, Hegel would argue that it is not only proven false, but also endures a destructive collapse into its own opposite, which he describes in the *Science of Logic* as a process of ‘zugrunde gehen’. For Hegel, freedom is *negated* by the LNC’s denial of contradiction, whereby it vanishes into and becomes the other. If Hegel’s dialectic did not preserve the LNC, there would be no movement. The dialectic requires a criterion in accordance with which one in a pair of contradictory opposites can be proven false in order for it to collapse into the other. The legitimate demonstration of the falsity of one in accordance with the LNC is what necessitates the movement to the other. So conceived, the LNC is not a mere formal law that regulates the correctness of an equally formal logical argument, but is *a negation constitutive of the movement and becoming* of the content itself, i.e. of the opposites themselves.

The collapse of one opposite into another does not constitute the genuinely dialectical – what Hegel also calls the *speculative* – nature of logic. This only manifests when the movement of collapse is doubled in the form of an antinomy. The genuinely dialectical is not constituted simply by the one-sided collapse of the thesis into the antithesis or by the equally one-sided collapse of the antithesis into the thesis, but is both movements taken as an antinomial whole. In that case, the contradictory opposites of thesis and antithesis demonstrate themselves to be the same *identical* collapse of each into the other through the mediating power of the LNC qua double negation. The thesis violates the LNC and, through this first negation, becomes the antithesis; at the same time, the antithesis also violates the LNC and, through this second negation, becomes the thesis. The result is the standpoint of the dialectic in which contradictory opposites constitute a single and utterly dynamic antinomial whole. We can now ascertain the Hegelian position that the analytic is not abandoned by, but preserved (*aufgehoben*) within the dialectic. The point that must be made clear is that the thesis and the antithesis of an antinomy, when not taken together as a whole, but as independent from one another, represent distinct analytic arguments. Hegel’s insight is that this same kind of analytic reasoning can be legitimately applied to contradictory opposites in the two-sided form of an antinomy; and that this doubling of the analytic is dialectical. The point is that Hegel’s dialectic does not overcome, but *is itself the analytic grasped in its antinomial truth*. The aim is not to abandon or reject the analytic, but to demonstrate how it develops into and becomes dialectical.

**Hegel’s Dialectic as Bridge**

A contemporary analytic philosopher who agrees, in principle, with this interpretation is Graham Priest. In his book *In Contradiction*, Priest also locates the main influence on Hegel’s dialectic in the Kantian antinomies; and

---

1Hegel, *Science of Logic*, p. 434
he similarly recognizes that what Hegel took from Kant is the insight that the contradiction displayed in an antinomy is the result of the analytic argumentation in both the thesis and the antithesis. Priest states:

In his *Logic*, Hegel agreed with Kant that the antinomies, the arguments that end in contradiction, proceed by perfectly legitimate reasoning. However, he found no basis for ruling the application of concepts within them to be illegitimate ... Thus, according to Hegel, perfectly correct reasoning, using legitimate application of certain concepts, leads to contradiction.¹

Priest is not interested in Hegel for historical, but for logical reasons. He refers to him in order to draw attention to a logical theory that he has developed called *dialetheism*. This title is derived from the term, *dialetheia*, which Priest defines as ‘any true statement of the form: $\alpha$ and it is not the case that $\alpha$’.² The main idea behind the theory is that there are acceptable cases of contradiction, i.e. *dialetheia*, in the common everyday usages of language as well as in the technical and complex language of symbolic logic. Although Priest does not agree with Hegel on the pervasiveness of contradiction, he acknowledges that, in certain respects, he is correct:

The only point that I wish to isolate and highlight is Hegel’s contention that our concepts are contradictory, that there are true contradictions... It is the main claim of this book that Hegel was right: our concepts, or some of them anyway, are inconsistent, and produce *dialetheias*.³

One of the compelling parts of Priest’s theory is the seriousness with which he addresses Hegel’s logic from an analytic point of view. This is rare indeed. Unfortunately, Priest does not, but certainly could use *dialetheism* to engage in the analytic/continental debate. Nevertheless, Priest corroborates the position of this work by likewise pinpointing the aspect of Hegel’s logic that could bridge these traditions – namely, that the dialectic is a logic of antinomies that proceeds on the basis of ‘perfectly correct reasoning.’ This may hopefully assuage the concerns of those who believe that Hegel discards basic principles of logic; he does not. More important than this, however, is that Priest is someone, from the analytic side, whose philosophical convictions confirm the possibility that Hegel’s dialectic provides a logical scheme that could reconcile what some, like Rockmore, claim is the intractable conflict between analytic and continental philosophy.

Rockmore’s incompatibility interpretation provides us with a good test case to see just how the dialectic would function as a bridge between the analytic and continental traditions. To recall, Rockmore contends that the chief reason

¹Priest, p. 4
²Ibid, p. 5
³Ibid, p. 4
for the incompatibility is that these traditions developed out of philosophically opposed responses to Kant’s idealism: the continental tradition affirmed the constructivism of his idealism as detailed in the Copernican revolution; whereas the analytic tradition rejected this in favor of a predominantly realist set of convictions. Although one may rightly be hesitant to accept the reduction of these highly complex and often internally conflicting traditions to the general distinction between idealism and realism, Rockmore must be given credit for capturing what is certainly one of the many underlying trends that have shaped and still continue to shape this conflict. If, therefore, we concede Rockmore’s point, then the incompatibility can be expressed in the form of an antinomy as follows: on the side of the thesis is the idealist assumption of the continental tradition that to know objects is to construct the conceptual framework for cognizing them as they occur within the mind; on the side of the antithesis is the realist assumption of the analytic tradition that to know objects is to uncover, discover, or reveal them in a world independent of and external to the mind.

This antinomy is not at all alien to Hegel in that it roughly follows along a similar line of reasoning found in the Inverted World of the *Phenomenology of Spirit*. Dialectically, the thesis and the antithesis are not mere formal arguments, though they can certainly be presented that way, but movements of zugrunde gehen in which each collapses into the other. This kind of movement, however, is only possible if the LNC is not abandoned by, but preserved within the dialectic. On the side of the thesis, realism collapses into idealism on the grounds that it violates the LNC; conversely, on the side of the antithesis, idealism collapses into realism on the same grounds that it also violates the LNC. For Hegel, the genuinely dialectical moment of this antinomy is not that each side contains its own independent movement, but that both sides simultaneously collapse into one another to form a single, dynamic antinomial whole. From the vantage point of the whole, the distinction between idealism and realism that Rockmore attributes to the continental and analytic traditions respectively is not a fixed opposition of incompatibles, but a fluid interaction of opposites in which each crosses over into the other. Admittedly, Hegel’s terminology can be off-putting, with the sometimes unfortunate consequence that the main idea is overshadowed by the novelty of expression. Nevertheless, the dialectical point is simply this: the bridge of reconciliation between the analytic and continental traditions is their mutual fallibility. Neither side has an absolute claim to philosophical supremacy over the other insofar as both violate the LNC. Again, to use Hegel’s at times taxing terminology, the bridge is constituted by the fact that analytic and continental philosophy, despite the antagonism of their differences, are also the same in that neither fully abides by the LNC. Expressed dialectically, analytic and continental philosophy are opposites that are also identical qua the movement of collapse of each into the

---

1 Carnap’s *Logischer Aufbau der Welt* could be used to counter Rockmore’s distinction here. In this text, Carnap seems to offer an analytic version of what Rockmore calls “constructivism.”
2 Hegel, *Phenomenology*, §§154-164
other. It is this dynamic identity of mutual fallibility that reconciles the two traditions.

This is not intended as a pessimistic conclusion. The point is not that both traditions should be cast aside as inadequate on the grounds that neither completely lives up to the standard of non-contradiction; nor is it the cynical conclusion that there is no truth. Quite the contrary, there are a number of positive outcomes here: firstly, one must keep in mind that the goal of a *reductio ad absurdum* is to demonstrate the truth of something. Thus, analytic and continental philosophy each possess a relative truth over against the other. Secondly, in the everyday context of academia, the dynamic movement described above manifests as the on-going dialogue and critical exchange between the members of the two traditions. Over time, each side puts forth, from the standpoint of its own relative truth, a set of positive theories based on an array of philosophical convictions; at the same time, each side also operates as a critical opposition that scrutinizes and tests the legitimacy of the other. Thirdly, and most importantly, philosophy itself can no longer be understood as either analytic or continental alone. Rather, from the dialectical vantage point of the whole, philosophy *is* the on-going dialogue and critical exchange between the two traditions. This is the case whether we choose to make it explicit or not. It can already be seen in Brandom and Rockmore, among others. What is significant about the work of these thinkers is not that one is absolutely right and the other absolutely wrong, but that each is participating in the dialogue and critical exchange that constitutes philosophy as it is today. Perhaps Plato was on to something; perhaps participation is the bridge between opposites.

**Bibliography**


Edmund Burke is famous for his empiricist and physiological account of aesthetic pleasure in the beautiful and delight in the sublime. Burke argues that aesthetic pleasure or taste cannot be explained without taking into account our most fundamental human interests: the feeling of the beautiful is grounded in our social nature and, more specifically, in our passions that are concerned with ‘the society of the sexes’ (Enquiry, I, 9, p. 41-42)\(^1\), and our delight in the sublime is rooted in our desire for self-preservation. His empiricist and physiological theory of aesthetic pleasure was heavily contested by his contemporaries and was only revived through the later Nietzsche’s ‘physiology of aesthetics’ and his fierce attack on Kant’s and Schopenhauer’s account of aesthetic disinterestedness. Kant’s account of aesthetic disinterestedness can be understood (I shall argue) as a critical response to Burke’s empiricist account\(^2\), but – pace Nietzsche’s harsh mockery of Kant’s view – aesthetic disinterestedness, as Kant analyses it, does not imply any rejection of the relevance of the senses and the body in aesthetic judging. On the contrary, Kant’s transcendental critique of aesthetic judgment is – at least to a certain extent – reconcilable with Burke’s somatic theory, but repudiates the latter’s empiricist identification of the agreeable and the beautiful. Furthermore, like Burke, Kant emphasizes the social nature of the aesthetic experience, but argues (rightly) that Burke cannot justify the universal validity claim inherent in judgments of taste.

I devote the first part of this paper to a discussion of Burke’s and Kant’s views of aesthetic pleasure, especially in the beautiful, in order to show that Kant’s view of the disinterestedness of aesthetic pleasure or liking

\(^{1}\)References to Burke’s Enquiry are to E. Burke, A Philosophical Enquiry into the Sublime and Beautiful, edited with an introduction and notes by James T. Boulton, London, Routledge, 2008. Roman numbers refer to the part, Arabic numbers refer to the section, followed by the page number.

\(^{2}\)I do not claim that Kant’s aesthetic theory in the Critique of Judgment is a response only to Burke’s views on the beautiful and the sublime. It is not only impossible to discuss the historical context of Kant’s third Critique in a single essay, but it would also be absurd to reduce Kant’s treatment of aesthetics to a response to only one author. Kant not only criticizes Edmund Burke, but also Francis Hutcheson, David Hume, Lord Kames, Alexander Gerard, Alexander Baumgarten, and several others.
(Wohlgefallen) can be interpreted as a critical response to Burke’s failure to distinguish properly between the beautiful and the agreeable. The second part is concerned with the social values of aesthetic judgment and experience. Although Kant (wrongly) holds that the universal communicability of aesthetic judgments logically follows from the disinterested character of the pleasure on which they are based, Kant’s emphasis on the a priori validity of judgments of beauty can be viewed, or so I argue, as a rebuttal of the kind of empiricist and physio-psychological arguments that Burke offers to justify the social nature of the experience of beauty.

Burke and Kant on Pleasure and Disinterestedness

On a Burkean view, aesthetic pleasure can occur in at least two distinct ways. Something can be positively or negatively pleasurable. Pleasure and pain are, Burke contends, no mere relations that could only exist in contrast to some previous state of mind: there are pleasures and pains ‘of a positive and independent nature’ (Enquiry, I, 4, p. 35) and the diminution or cessation of pain does not result in positive pleasure, but in, what Burke calls, delight. Delight is related to privation, i.e., it is a pleasure ‘which cannot exist without a relation … to pain’ (Enquiry, I, 4, p. 36). The beautiful is the aesthetic variant of positive pleasure, whereas our feeling of the sublime is based on relative pleasure, i.e. so-called delight. Our delight in the sublime – ‘the strongest emotion which the mind is capable of feeling’ (Enquiry, I, 7, p. 39) – belongs to the passions of self-preservation. The sublime delight can arise only when there is danger and terror is felt: ‘A mode of terror, or of pain, is always the cause of the sublime’ (Enquiry, IV, 8, p. 134). The emotion of terror is closely related to privation of some sort: solitude as the privation of society, silence as the privation of sound, darkness as the privation of light. The feeling of the sublime occurs when this privation is suspended, and the fear or terror is postponed. We are threatened by loss: language, light, sound, life, everything threatens to disappear, and then, this terror of nothingness, this feeling of losing everything is suspended, and we experience delight. We experience the delight of being deprived of those privations. No moral catharsis occurs, as Aristotle thought. The delight in the sublime offers no moral purification or elevation, but intensifies our affective capacities, and heightens our sensitivity.

Whereas the sublime is bound up with our sense for self-preservation and our fear of losing our capacities to live our own lives, the beautiful is a positive pleasure that is grounded in our social capacities and our desire to live with others. Not surprisingly, Burke connects the beautiful with love, which is ‘that satisfaction which arises to the mind upon contemplating anything beautiful’ and which needs to be distinguished from desire or lust, ‘which is an energy of the mind, that hurries us on to the possession of certain objects, that do not affect us as they are beautiful, but by means altogether different.’ (Enquiry, III, 1, 91) Beauty is a social quality, ‘for where women and men, and not only they, but when animals give us a sense of joy and pleasure in beholding them,
The Social Nature of Beauty: Kant’s Response to Burke

(and there are many that do so) they inspire us with sentiments of tenderness and affection towards their persons; we like to have them near us, and we enter willingly into a kind of relation with them, unless we should have strong reasons to the contrary.’ (Enquiry, I, 10, p. 42-43)

Kant not only reacts against the rationalists who wrongly ‘intellectualize’ aesthetic experience by assimilating the beautiful to the good, but also attacks the advocates of an empiricist and physiological approach, and especially Burke, since Kant says that he ‘deserves to be named as the foremost author in this sort of approach’ (CJ, 5: 277). The first and most obvious reason for this repudiation of Burke’s ‘physiological exposition’ (ibid.) is that it cannot properly distinguish between the feelings of the agreeable and the beautiful. On the physiological view, the difference is merely a difference in degree and not in quality. The second is that this approach cannot account for, what Kant calls, the ‘pluralistic’ nature of aesthetic judgments (CJ, 5: 278), i.e. the idea – which Kant shares with rationalist predecessors such as Mendelssohn and Baumgarten – that in matters of aesthetic taste, there is a genuine ‘reason to have controversy about taste, not merely to shrug one’s shoulders and say “to each his own”: because judgments of taste rest upon some sort of judgment of the object, specifically of the object’s form’. Thus, contra Hume, Burke and other empiricists, Kant argues that aesthetic judgments justifiably make claims to universal validity. We value beauty not just guided by our own private or ‘egoistic’ interests (CJ, 5: 278), and nor because (as Burke holds) because beauty stimulates our social passions, such as love. We experience and appreciate beauty as a priori shareable with others who possess similar discriminatory and judgmental capacities. I shall return to this second issue in the second part of my paper, and now concentrate on the first one: the distinction between the agreeable and the beautiful, and why Kant believes that disinterestedness is a suitable criterion to distinguish qualitatively between both feelings.

Although Kant agrees with Burke that, ‘as Epicurus maintained, gratification and pain are always ultimately corporeal … because life without the feeling of the corporeal organ is merely consciousness of one’s own existence, but not a feeling of well- or ill-being’, and praises Burke’s analyses of (aesthetic) pleasure and displeasure as ‘extremely fine’ and admits that his ‘psychological remarks … provide rich materials for the favorite researches of empirical anthropology’ (CJ, 5: 277), he rejects his empiricist assimilation of pleasure in the beautiful to merely agreeable sensation.

How does Kant distinguish the feeling of the beautiful from the agreeable? Pleasure in the agreeable is, Kant argues, ‘interested’. There is much debate in the literature about the exact meaning of this phrase. As Nick Zangwill rightly

---

1The abbreviation CJ refers to Immanuel Kant, Critique of Judgment. Citations to the Critique of Judgment are to volume 5 and the section and page numbers of the Akademie-Ausgabe (Berlin, de Gruyter, 1902-). The English translations are all based on I. Kant, Critique of Judgment, translated, with an introduction, by Werner S. Pluhar, Indianapolis, Hackett, 1987. I have modified this translation where it seemed appropriate.

remarks, ‘many commentators have found Kant’s account problematic if not completely unintelligible.’ Whereas I do not pretend to be able to completely clarify this complex notion, we do need to linger on it for a while here, for the question of (dis)interest is crucial to a proper understanding of Kant’s qualms about Burke’s physiological approach, which (according to Kant) unjustly identifies the pleasure in the beautiful with the pleasure in the agreeable. Kant argues that pleasure in the beautiful is disinterested, unlike our pleasure in the agreeable. In section 2 of the Critique of Judgment, Kant writes that ‘the satisfaction that we combine with the representation of the existence of an object is called interest. Hence such a satisfaction always has at the same time a relation to the faculty of desire, either as its determining ground or else as necessarily interconnected with its determining ground.’ (CJ, 5: 204) In his insightful paper on ‘Kant on Pleasure in the Agreeable’, Zangwill clarifies this as follows: ‘if a pleasure is an “interest”, in Kant’s sense, it means that it bears an intimate relation to a desire (that is, a concern with real existence). An “interest” is a pleasure that has some kind of necessary connection with desire. A pleasure is “disinterested” if it has no such necessary connection with desire’. It is worth noting that Kant’s conception of interest is broader than the idea of self-interest which Kant’s reference to the capacity of desire (Begehrungsvermögen) seems to suggest: pleasures in the good are as ‘interested’ as pleasures in the agreeable. By contrast, Kant holds that our pleasure in the beautiful cannot originate from any interest, but also, and perhaps more importantly, that our pleasure in the beautiful does not create any interest in the object either.

What can this mean? As Kant asserts at the beginning of § 5, ‘a judgment of taste is merely contemplative, i.e., it is a judgment that is indifferent with regard to the existence of the object: it considers the character of the object only by holding it up to the feeling of pleasure and displeasure (nur seine Beschaffenheit mit dem Gefühl der Lust und Unlust zusammenhält)’ (CJ, § 5, 5: 209). Agreeable objects, say Belgian chocolates, which cause pleasure merely because of their sensible properties, are said to ‘gratify (vergnügen)’ someone. More specifically, when I enjoy eating a Belgian chocolate, ‘I am not granting mere approval: the agreeable produces an inclination’ and ‘arouses a desire for objects of the same kind’ (CJ, § 3, 5: 207). Thus Kant holds that ‘all interest presupposes a need or gives rise to one; and, because interest is the basis that determines approval, it makes the judgment about the object unfree.’ (CJ, § 5, 5: 210) Hence, the basic difference between the agreeable and the beautiful would be that the agreeable gives rise to a desire for similar objects, whereas the feeling of the beautiful does not. Kant thus plausibly argues that pleasure in

---

3 In ‘Kant on Pleasure in the Agreeable’, Zangwill rightly emphasizes that this unfreedom of the pleasure in the agreeable is ‘a matter of the causes of the pleasure. It does not detract from what Kant is saying about the way that pleasure then provokes desire, via a representation. If a pleasure is unfree, it is unfree because of the way it is caused, not because of what it causes.’ (p. 170)
The agreeable is connected with the \textit{existence} of the object that caused the agreeable sensation in the first place, whereas pleasure in the beautiful is not. If the satisfaction caused by the object leads to a desire for more similar objects, e.g. more Belgian chocolates, then this implies that the initial satisfaction was connected with the existence of the first object. How else could it produce this desire (or inclination) for more objects that are thought to be similar?

Kant’s analysis of pleasure in the agreeable raises a number of worries. First, how plausible is Kant’s claim that pleasures in the agreeable are necessarily productive of desire for more similar objects? Not all pleasures in the agreeable seem to provoke the desire for more objects of the same kind. Put more concretely, as Zangwill asks, ‘what about the last piece of chocolate that we enjoy before we have had enough? … The sight of yet more chocolate can soon come to disgust one. It seems that the last pleasurable piece of chocolate does not provoke a desire for more of the same’. Thus, although Kant may be right that many kinds of agreeable sensations are ‘more-ish’ or productively interested, not all pleasures in the agreeable are. There is a second possible objection to Kant’s account, viz. that Kant, as Paul Guyer notes, instead of distinguishing between kinds of pleasure, merely supplies ‘a distinction between feelings of pleasure and all other kinds of sensation’. However, Kant’s view of pleasure is more complicated than Guyer allows. By defining pleasure as \textit{feeling} instead of sensation, Kant is not merely saying that pleasure is some peculiar kind of sensation, i.e. a subjective sensation \textit{which cannot become an element of cognition at all} (\textit{CJ}, Intro., VII, 5:189; bold in the original). The subjective nature of Kant’s notion of feeling is much more profound than that. Rachel Zuckert suggests (rightly) that ‘pleasure is, on Kant’s definition, a representation with intentional content, which comprises other representations understood to be modifications of the subject (that is, are themselves not [solely] referred to objects). In the \textit{Critique of Judgment}, Kant defines pleasure as the ‘consciousness of the causality of a representation with respect to the state of the subject, \textit{for maintaining} it in that state’ (\textit{CJ}, § 10, 5:220; bold in the original) and in the \textit{Anthropology from a Pragmatic Point of View}, he contends that ‘what directly (through sense) urges me to \textit{leave my

\footnotesize

\begin{enumerate}
\item This does not necessarily imply that Kant is offering a purely causal account of the interestedness of pleasure in the agreeable. I here agree with Zangwill, ‘Kant on Pleasure in the Agreeable’, p. 169: ‘Once we see that Kant is not offering a purely causal account of the interestedness of pleasure in the agreeable, we will be less prone to think that he thinks that pleasure in the beautiful is disinterested because the pleasure bears no causal relation to the objects that we find pleasurable and thus call beautiful. If Kant did think this, it would make his claim that pleasure in the beautiful is disinterested very implausible. But fortunately Kant holds no such view.’
\item Zangwill, ‘Kant on Pleasure in the Agreeable’, p. 172.
\item Guyer, \textit{Kant and the Claims of Taste}, p. 153.
\item Zuckert, \textit{Kant on Beauty and Biology}, p. 233. I here follow Zuckert’s excellent account (pp. 233 ff.) of the intentional nature of pleasure, but I do not concur with her identification of the intentionality of pleasure with purposiveness without a purpose.
\item Zuckert, \textit{Kant on Beauty and Biology}, p. 233.
\end{enumerate}
state (to go out of it) is disagreeable to me – it causes me pain; just as what drives me to maintain my state (to remain in it) is agreeable to me, I enjoy it.’ (Anth., 7: 231; italics in the original)1 Thus Kant does not agree with Burke (and other empiricists) that pleasure is a kind of primitive or raw sensation, but holds that pleasure is a feeling about something, or more accurately, about the continuation of the feeling or the mental state.2 Pleasure in eating Belgian chocolates would then be the awareness or ‘the feeling that the representation of chocolate is “causing” one to stay in the state of having that representation (of the taste of chocolate).’3 Pleasure is thus intimately connected with the feeling of life (Lebensgefühl) (CJ, § 1, 5: 204; 277), i.e., with enjoying the state one finds oneself in when (for instance) experiencing the sensible properties of an object. Thus, on a Kantian view, pleasure is no mere ‘raw feel’, as Guyer, along with numerous other commentators, claims. It does not need to be referred to objects via empirical concepts or judgments, but is necessarily characterized by intentionality, i.e., ‘aboutness’: it ‘is about’ a subject’s mental state. Therefore, it is aptly called subjective by Kant, although it is not a sensation, but ‘a second-order, reflexive state with respect both to other mental states and to the position of those states in time, the form of inner sense.’4 We do not experience pleasure primarily as the separate effect of something, but we take pleasure in something (ibid.), e.g., in drinking a glass of Chablis, in eating spinach, in sinking into a hot bath, etc.

Contra Burke, and although bodily pleasures such as a sexual orgasm or tasting a fine wine may seem to suggest otherwise, Kant suggests that pleasures are not free floating sensations but reflexive, second-order feelings. Contrary also to the earlier view defended in the Critique of Practical Reason, in the Critique of Judgment he now claims that not all pleasures are sensations or, more accurately, sensory pleasures are, pace Burke, not the only kind of pleasure – although he does retain the view that when pleasures are ‘sensations’ (pleasures in the agreeable or ‘enjoyments’), they are ‘the same in kind, differing only in degree’.5 Of course, he agrees with Burke that sensory pleasure is a kind of pleasure, but refuses to accept Burke’s privileging of sensory (or bodily) pleasure – the sensuous pleasure we take in enjoying a cognac or a hot shower – as a model for all other kinds of pleasure. Hence, Kant maintains that pleasures in the agreeable are only a sub-class of pleasure; they are the kind of sensory pleasures that we share with animals (CJ, § 5, 5: 210). There is no reason, though, to privilege these agreeable sensations over other kinds of pleasure.

As previously noted, Kant rashly claims that all pleasures in the agreeable are productively interested, i.e. arouse the desire for more objects of the same kind. ‘Sated’ pleasures, though, such as orgasms, do not – at least not

---

1I here refer to I. Kant, Anthropology from a Pragmatic Point of View, translated by Mary Gregor, The Hague, Martinus Nijhoff, 1974.
2See Zuckert, Kant on Beauty and Biology, p. 233.
3Zuckert, Kant on Beauty and Biology, p. 233.
4Zuckert, Kant on Beauty and Biology, p. 236.
5Zuckert, Kant on Beauty and Biology, p. 240.
immediately – provoke the desire for more similar pleasures. But what about Kant’s insistence that pleasure in the beautiful is devoid of interest, i.e., is only related to the subject’s feeling of life (Lebensgefühl), and is completely independent of the existence of the object? Kant argues that pure aesthetic pleasure is directed to the representation of the object, as opposed to the connection between the subject and the existence of the object. Kant’s basic idea is that of the contrast between the mere representation of an object and the full nexus of its causal relations. Only in the case of the latter can we have empirical knowledge of its causal basis. A physiological response can be the subject of empirical investigation and empirical causal laws. The agreeableness of the object may be included in the causal nexus that constitutes the real existence of the object, whereas the feeling of the beautiful cannot – again pace Burke, who claims that ‘the appearance of beauty as effectually causes some degree of love in us, as the application of ice or fire produces the ideas of heat or cold’ (*Enquiry*, III, 2, p. 92).

From a Burkean perspective, the beautiful causes the passion of love. Although Burke distinguishes love from desire, and (only) in this way anticipates Kant’s analysis of disinterestedness, he offers no solid basis to explain the qualitative distinction between the beautiful and the agreeable. As an empiricist, he can only account for a difference in degree. Kant holds not merely that Burke’s distinction between desire and love is flawed, but also that his physiological explanation of the beautiful cannot account for the ‘pluralistic’ character of judgments of beauty. It is to this issue that we now turn.

**The ‘Universal Voice’ of Beauty**

Although Burke emphasizes the social nature of beauty and love several times, he remains silent on the question of the universality claim of judgments of beauty. He does connect beauty with our ‘passions for society’, but this has more to do with the passion caused by the experience of beauty, viz. love, than with the appreciation of beauty as such. Kant severely condemns any attempt to dispense with the objectivity claim of pure aesthetic judgments and rejects Burke’s contention that beauty is derived solely from sensations that depend merely on our physiological constitution. When we judge something to be agreeable, Kant says, we can accept that others disagree: ‘this dish is agreeable to me’ is an acceptable expression (*CJ*, § 7, 5: 212). But when I judge something as beautiful, I cannot claim that it is merely beautiful to me (*CJ*, § 7, 5: 212): although the judgment is based on a personal feeling of pleasure (Wohlgefallen), we require or demand others to agree with us: a judgment on the beauty of an object is always pluralistic.

In some ways, Kant’s view is even closer to Burke’s than to other empiricist views such as Hume’s, for Burke holds that the principles of taste are uniform, whereas Hume’s famous essay ‘Of the Standard of Taste’ emphasizes the great

---

variety between people’s judgments of taste and argues that the only real standard of taste to be found has to be based on an historical canon and the joint verdict of an elite of trained critics, which is the most reliable standard we can possibly have. Burke, however, urges that: ‘as the senses are the great originals of all our ideas, and consequently of all our pleasures, if they are not uncertain and arbitrary, the whole ground-work of Taste is common to all, and therefore there is a sufficient foundation for a conclusive reasoning on these matters.’ (Enquiry, p. 24) Unfortunately, he does not substantiate this rather bold claim.

Although Kant concurs with Burke’s (typically empiricist) emphasis on the role of the senses in matters of taste and stresses the importance of a personal encounter with the aesthetic object, he reprehends his conclusion that our common human physiology sufficiently grounds the universality claim of pure judgments of taste. Kant attempts to provide a priori foundations for what Burke thought were matters of natural principles, imagination, custom and physiological disposition. Contra Burke, who holds that ‘beauty is, for the greater part, some quality in bodies, acting mechanically upon the human mind, by the intervention of the senses’ Enquiry, III, 12, p. 112), Kant argues that even when, ‘as experience teaches’, a judgment of taste ‘is often enough rejected’ by others, we should not be deterred from demanding that others assent to it (CJ, § 7, 5: 213; § 8, 5: 214). However, Burke would never deny what Kant is claiming here, namely that the validity of a judgment of taste depends on the circumstances in which it is made. We may occasionally be mistaken that our own judgment of taste is not based on any personal interest (CJ, § 8, 5: 216; § 19, 5: 237), and we can only claim that others will judge the object in the same way, if the circumstances are ideal. Yet Kant still moves one (big) step further than Burke (and other empiricists), when he claims that we also make a claim ‘to everyone’s assent, as if it were an objective judgment’ (CJ, § 32, 2: 281; bold in the original) Here Kant clearly breaks with Burke, for he maintains that judgments of beauty are a priori rather than merely empirical.1

When we are judging something as beautiful, it is as if we speak with a ‘universal voice’, Kant says (CJ, § 8, 5:216). This universal voice is however not empirical, but ‘only an Idea’ in the Kantian sense of the term, i.e. a transcendental Idea to which no empirical representation conforms (CJ, § 8, 5: 216). Against the rationalists’ view, Kant insists that the required universal agreement in aesthetic judgments is always uncertain. One can reasonably claim that everyone should give his approval, but this requirement is not based on concepts, as rationalist philosophers argue to hold open the possibility for an ideal agreement or consensus. The evidence for (or against) my making a

---

1See also CJ, § 37, 5: 289: ‘And yet, that such judgments, even though the predicate (of one’s own pleasure that is connected with the presentation) is empirical, are nevertheless, as regards the demand that everyone assent, a priori judgments, or would be taken as such, is already implicit in the expressions used to make that claim; and thus this problem of the critique of judgment is part of the general problem of transcendental philosophy: How are synthetic a priori judgments possible?’ (Bold in the original.)
pure judgment of taste is uncertain, and it is not necessarily defeated by
disagreement either, because I might have been wrong about the source of my
own pleasure or because another may have not obtained the requisite
abstraction. It is founded on the idea of the harmony of our cognitive powers –
which is what Kant argues in § 9.

This is a crucial point: the claim to universal validity can neither be falsified
inductively, but nor – and here again Kant disagrees with Burke – can it be
verified empirically by basing one’s own judgment on the occurrence of (a
consensus of) other judgments of taste. Kant thus concurs with Burke’s claim
that judgments of beauty cannot be based on the subsumption of an object
under a determinate concept (such as perfection, as the rationalists hold). Thus,
if Kant is right that there is a claim to universal validity in pure judgments of
taste, the universality at stake is subjective:

[I]t does not contain an objective quantity of judgment, but only a
subjective one. For this quantity, I use the expression common
validity [Gemeingültigkeit], by which I mean the validity that a
presentation’s reference to the feeling of pleasure and displeasure
[may] have for every subject, rather than the validity of a
presentation’s reference to the power of cognition. (We may,
alternatively, use the same expression, universal validity, for both
the aesthetic and the logical quantity of a judgment, provided only
that we add objective for the logical universal validity, to distinguish
it from the merely subjective one, which is always aesthetic.) (CJ, §
8, 5: 214-215)

This subjective universality of judgments of taste has to do neither with any
moral interest, nor with the content of the judgment but clearly with the
epistemic status of the judgment: its extension is not (as in a logical judgment)
a class of objects but ‘a class of possible human judges’.¹ This universal
validity cannot be based on the classification of the object under a concept: the
step from ‘This rose is beautiful’ to ‘All roses are beautiful’ is not guaranteed
by the first judgment’s universal validity (CJ, § 8, 5: 215). Singularity and
universality seem to be tied together in a pure judgment of taste: through a
singular judgment, i.e. in confrontation with an object, the universal
shareability of the feeling of pleasure is claimed, without any reference to
determinate concepts.

Phenomenologically speaking, the purity of taste – the disinterestedness of
the experienced pleasure – is subjectively determined (as its subjective
condition, Kant says) by the necessity of being universally communicable or
shareable, i.e., by the ‘signal’ in the mind (Gemüth) that the felt pleasure (or
displeasure) is universally communicable. This implies that, on the one hand,
the disinterestedness of the pleasure is the essential, a priori condition for the
universal communicability of aesthetic judgments – without disinterestedness
aesthetic judgments could not be universally shared, for disinterestedness, or so

¹P. Guyer, Kant and the Claims of Taste, p. 132.
Kant argues, implies the purposeful ‘play’ of the cognitive powers. But, on the other hand, the capacity for universal communicability is itself the ideal gauge to estimate whether the experienced pleasure is really disinterested or not. Thus, Kant writes that:

If the pleasure in the given object came first, and our judgment of taste were to attribute only the pleasure’s universal communicability to the presentation of the object, then this procedure would be self-contradictory. For that kind of pleasure would be none other than mere agreeableness in the sensation, so that by its very nature it could have only private validity, since it would depend directly on the presentation through which the object is given. Hence it must be the capacity for being universally communicated of the mental state \[\text{allgemeine Mittheilungsfähigkeit des Gemüthszustandes}\], in the given presentation, which underlies the judgment of taste as its subjective condition, and the pleasure in the object must be its consequence. (\textit{CJ}, § 9, 5: 217; bold in the original)

Only through the possibility of universal communication can it be estimated, Kant holds, whether the felt pleasure is actually disinterested or not. So in this sense, and in this sense only, can the pleasure in an object be the consequence of ‘the capacity for being universally communicated of the mental state \[\text{allgemeine Mittheilungsfähigkeit des Gemüthszustandes}\]’. Whether or not the pleasure is really pure pleasure, i.e., the disinterested pleasure that grounds a pure judgment of taste, depends on the very universal communicability of the aesthetic judgment, which is the \textit{ratio cognoscendi} of the disinterestedness of the pleasure.\footnote{For a more extended discussion of this, see B. Vandenabeele, ‘The Subjective Universality of Aesthetic Judgements Revisited’, \textit{British Journal of Aesthetics}, 48 (2008), p. 410–425.} That one could actually be fairly certain (though one will never be able to prove it by means of arguments) that the pleasure one experiences here and now is disinterested, is grounded in – though not caused by – the universal communicability of the mental state – or more precisely still, in the \textit{affect} that ‘signals’ whether or not the activity of the mental powers is universally communicable. And this affect, or rather this universally communicable state of mind, of course, presupposes ‘a capacity for being universally communicated’.

\textbf{Conclusion}

A number of different traits run through Kant’s responses to the Burkean form of physiological and empiricist analysis of aesthetic pleasure that he diagnoses. One key feature of his approach is to argue that the empiricist method cannot account for the qualitative difference between the agreeable and the beautiful, since it does not acknowledge his (controversial) criterion of the disinterestedness of the pleasure on which a judgment of beauty is based.
Another characteristic of his approach is essentially to argue that Burke’s empirical psycho-physiological analysis of beauty should ultimately be refuted, since it cannot account for the aesthetic judgment’s claim to universal assent, and hence fails to adequately describe the extremely significant ‘pluralistic’ or social nature of the feeling and judgment of beauty.

These few reflections are hardly meant to constitute an adequate assessment of Kant’s response to Burke’s aesthetics of the beautiful. My purposes in this essay have only been to show that Kant is profoundly concerned with Burke’s empiricism, and that recognizing that his transcendental critique of the judgment of beauty is, to a certain extent, structured around the task of responding to this Burkean form of physiological empiricism may be a useful way to analyze his approach to aesthetic pleasure and appraisal.
In *Idea for a Universal History* and *Perpetual Peace* Kant elaborates a historical-teleological view on the development of a cosmopolitan political community oriented to realising the highest political good, i.e., perpetual peace. In *Religion within the Boundaries of Mere Reason* he exposes an analogous account on the establishment of a cosmopolitan ethical community oriented to realising the highest moral good, i.e., the union of virtue and happiness. In these works, Kant not only points at similarities and differences between these two kinds of communities, but also suggests some interactions in the historical-teleological sequence of establishing these communities. In the *Religion*, Kant holds that an ethical community cannot be brought into existence without ‘the foundation of a political community’ (R 6:94).\(^1\) Hence, the establishment of a political community seems to be a prerequisite for the establishment of an ethical community. Furthermore, Kant relates the establishment of these communities to the establishment of perpetual peace in two different manners. First, following *Idea for a Universal History* and other works, the formation of a political community leads to the establishment of perpetual peace, which in turn anticipates the realisation of an ethical community. Secondly, following passages from the *Religion* and *Perpetual Peace*, the establishment of an ethical community assures the realisation of perpetual peace. Taken together, these views imply that a political community leads to peace, that peace anticipates an ethical community, and that an ethical community once again assures peace.

The central question of this paper therefore concerns whether these views are as perplexing as they appear to be, or whether they can be meaningfully reconciled. We will claim that Kant’s views can be made compatible by distinguishing *legal* from *moral* peace, and argue for one historical-teleological pathway that subsequently leads from a cosmopolitan political community to a legal but *provisional* peace, from a legal peace to a universal ethical

---

An Anthology of Philosophical Studies

community, and finally from an ethical community to a moral and lasting peace. We will also claim that this non-circular pathway is related to a dynamical movement from morality to politics and vice versa. Before unfolding our interpretation, we will first outline the main features, analogies and differences of Kant’s ethical and political communities as well as Kant’s historical-teleological views on the establishment of these communities.

Kant’s Views on Ethical and Political Communities

In the *Groundwork of the Metaphysics of Morals* and *Religion within the Boundaries of Mere Reason* Kant introduces the concepts of a realm of ends (*Reich der Zwecke*) and an ethical community (*ethischesgemeines Wesen*). Kant describes the realm of ends as ‘a whole of all ends in systematic connection (a whole both of rational beings as ends in themselves and the ends of his own that each may set himself)’ (G 4:433). The realm of ends concerns an ideal of reason with which to compare our human relations and interactions: it serves for designing a vivid image of what would be a perfect social moral practice in order to realise one’s moral existence by the free submission to one’s own self-legislation and the awareness of the higher vocation of humanity. Kant therefore distinguishes the realm of ends from the realm of nature: whereas the latter concerns ‘a theoretical idea for explaining what exists’, the former concerns ‘a practical idea for the sake of bringing about, in conformity with this very idea, that which does not exist but which can become real by means of our conduct’ (G 4:436). Further, as equal moral self-legislators, all members of the realm of ends have to treat themselves and each other always as ends in themselves and never as mere means to ends. Hence, one obtains not only a union of rational lawgivers as ends in themselves, but also a non-conflicting unity of their common substantive purposes amongst a plurality of their particular relative purposes. Guyer (2002:§9) thus speaks of the realm of ends as ‘a system of freedom, in which all agents freely pursue their freely chosen ends to the extent compatible with a like freedom for all’. This freedom is not to be seen as absolute or unrestrained freedom, but rather, as responsible freedom that takes into account others’ points of view, that values humanity as an end in itself, and that only exists within an intersubjective community of equal self-legislators. Kant’s realm of ends is thus characterised by autonomous legislation, intersubjective reciprocity, and mutual responsibility.¹

In the *Religion* Kant elaborates this idea in terms of an ethical community: ‘An association of human beings merely under the laws of virtue, ruled by this idea, can be called an *ethical* and, so far as these laws are public, an *ethico-civil* (in contrast to a *juridico-civil*) society, or an *ethical community*.’ (R 6:94–95). Kant argues for both analogies and differences between an ethical community

---

Kant on Political and Ethical Communities with Regard to Perpetual Peace

and a political community. First, Kant claims that an ethical community cannot be equated with a juridical commonwealth or a political state. The latter aims at protecting the individual’s rights by means of legal force: it stipulates which actions are legal, and it safeguards the individual’s right to freedom by exercising public and external coercion. Hence, it is authorised to force the individual to restrict his actions so that these agree with the same right to freedom of all other individuals (R 6:98), but it cannot determine the ends set by man – insofar as these ends do not lead to juridical offences – for the setting of ends implies an act of freedom and such freedom has to be warranted by jurisdiction. In contrast to a juridical-civilian community, the membership in an ethical-civilian community has to be based upon free consent instead of external coercion. Thus, whereas a juridical-civilian community is characterised by the interaction of individuals insofar these are collectively submitted to public, juridical and coercive laws, an ethical-civilian community is characterised by the union of people under laws which are free from coercion, i.e., laws of virtue (R 6:95). Secondly, since an ethical community aims at uniting all subjects into one system of ends, it cannot be susceptible to any restriction concerning its range and extent. In contrast to a political state, which is traditionally limited to a territorially delineated population, an ethical community is essentially cosmopolitan. Analogously, an ethical community cannot be bound by any kind of specific habits, customs, traditions, or beliefs on the basis of which certain groups can be denied their membership. An ethical community is thus quantitatively characterised by its universality (R 6:101). Thirdly, an ethical community can only acknowledge moral purposes and motives – i.e., purposes and motives derived from reason based, freely accepted and universally binding moral laws – as legitimate ones. An ethical community is thus qualitatively characterised by the moral purity of one’s motivation for membership in this community (R 6:101). Finally, the constitution of an ethical community, i.e., the leading principle of its union and unity, has to be invariable with the exception of the contingent rules that relate to the administration of this community and that have to be adapted to changing (time)circumstances (R 6:102). At the same time, this constitution has to be flexible and modifiable in order to reflect the free rational judgment of its members and to comprise the whole of humanity in a constant progress for the ethical community’s membership necessarily stretches to all human beings.¹

Although disagreements remain concerning the details of Kant’s conception of a cosmopolitan political community, it suffices for our purposes to outline the following features. In dealing with perpetual peace as the highest aim of such a community, Kant distinguishes three parts of public right that come into play: the rightful regulation of the interactions among the state’s individuals by means of rules of law, the rightful regulation of the interactions among states by means of rules of international right, and the rightful regulation of the interactions between states and foreign individuals by means of rules of international right.

¹For further analyses of Kant’s ethical community, see Anderson-Gold, 2001:33-52; and Michalson, 2001.
cosmopolitan right. These three parts of public right are mirrored in Kant’s three definitive articles for establishing international peace: ‘The civil constitution in every state shall be republican’ (PP 8:349), ‘The right of nations shall be based on a federalism of free states’ (PP 8:353), and ‘The cosmopolitan right shall be restrained to the conditions of general hospitality’ (PP 8:356). All three parts of public right are grounded in the underlying idea of external freedom: the notion of right is derived from the concept of freedom applied to the external relations among individuals. Kant thus defines ‘right’ as ‘the limitation of the freedom of each to the condition of its harmony with the freedom of everyone insofar as this is possible in accordance with a universal law; and public right is the sum of external laws which make such a thoroughgoing harmony possible’ (TP 8:289–290; MM 6:230). For Kant a cosmopolitan political community – organised as a republic – is the only system that fully endorses the requirements of public right and is characterised by the external freedom of all members of society as humans, their legal equality among each other as civilians, their independence as legal co-legislators, and their dependence on a common juridical legislation (PP 8:349; TP 8:290–294).¹

Bearing the discussed features in mind, let us turn to Kant’s historical-teleological views on ethical and political communities. In Idea for a Universal History Kant argues that the absence of empirical evidence concerning the moral progress of the human species cannot discourage us from reflecting on this issue on behalf of our moral vocation, and to formulate some ‘suppositions’ concerning how nature could contribute to such progress (I 8:29–31). Kant claims that the purpose prescribed by nature on behalf of humanity exists in the formation of a ‘civil society’ in order to protect the freedom of every individual (I 8:22), and identifies nature’s driving mechanism as the unsociable sociability of man: on the one hand, this outer natural disposition urges mankind to enter into political communities out of self-interest, while, on the other hand, it drives mankind to undermine such communities by striving for supremacy over the fellow-man and by curtailing his right to freedom. Such antagonism inevitably leads to the erection of political states where the lawless will of every subject becomes subordinated to a political authority and where everyone is forced to obey universally valid (juridical) laws so that the external freedom of everyone can be guaranteed (I 8:23). However, because political authorities might place themselves above the laws and violate the civilians’ rights, Kant argues that the final political task of mankind exists in the inauguration of a constitution so that state-power can be exercised in an adept way with regard to the civilians, and in the construction of an international legislation so that the mutual relations between states can be regulated.² In Perpetual Peace Kant further argues that external freedom and right can be fully realised only in a cosmopolitan union of states under common international laws: only as such, can man can enter the era of perpetual peace. Although Kant struggles with the exact determination of a

¹For a commentary hereof, see Rauscher, 2007:§2.
²See also Kersting, 1992:361-362.
league of nations or a cosmopolitan (con)federation,¹ and although he expresses his doubts concerning the actual realisation of a peace-loving world union, he never backs down from approaching the idea of perpetual peace as a necessary condition for realising a moral politics of international right and stresses the endless approximation of perpetual peace as a moral duty (MM 6:350). Hence, Kant describes the process of peace-making as the ‘final end of the doctrine of right’, and calls perpetual peace the ‘highest political good’ (MM 6:355).

Taking into account these views on the development of a cosmopolitan league of nations oriented to perpetual peace as the highest political good, in the Religion Kant clearly develops a likewise historical-teleological view regarding the realisation of an ethical community oriented to the union of virtue and happiness as the highest moral good. Wood (1992:409-410) sums up the following analogies. First, just as states can be regarded as empirical representations of a universal political community or a realm of external legal freedom, likewise churches embody empirical forms of a cosmopolitan ethical community or a realm of inner moral freedom. Secondly, just as states have often shown shortcomings in their task of realising right, likewise churches have all too often neglected their task of striving for true morality. Thirdly, just as states have to be conceived as transitional stadiums towards a cosmopolitan league of nations, and hence have to be assigned with the historical function of reorganising themselves into a political community that universally guarantees right and external freedom, likewise ecclesiastic faith functions as a vehicle for moral religion so that churches have to reform themselves into an ethical community that aims at the universal realisation of virtue and inner freedom. Additionally, Rossi (2005:§3.8) argues that both political and ethical communities are to be seen as ‘factors in forming the trajectory of human history toward what Kant sees as the moral destiny of humanity as a species’.

Results and Discussion. Interpreting the Historical-teleological Sequence of Establishing Ethical and Political Communities with Regard to Perpetual Peace

Bearing the above considerations in mind, let us discuss Kant’s different views regarding the historical-teleological sequence of establishing these communities and realising perpetual peace. Kant’s first view entails that the establishment of a political community realises peace and, as such, anticipates the establishment of an ethical community. We call this view the standard view as it is sustained by extensive textual evidence. In Idea for a Universal History Kant argues that the formation of a political community as a prerequisite for perpetual peace not only involves the termination of violence, but also leads to the spending of government funding for the elaboration of social provisions and the promotion of instruction, culture and education (I 8:26-28). Kant holds that precisely this triad promotes the transformation of a pathologically enforced political society into a ‘moral whole’, i.e., an ethical

¹Compare MM 6:350-351; PP 8:354; PP 8:356; PP 8:357.
community (I 8:21). Likewise, in Perpetual Peace Kant claims that because of the political guarantee for free opinion of speech, man will be able to free himself from the chains of tutelage, and hence to bring about the era of Enlightenment wherein morality can be fully accomplished. Furthermore, in the Religion Kant explicitly states that ‘without the foundation of a political community, [an ethical community] could never be brought into existence by human beings’ (R 6:94–95). Finally, in The Conflict of the Faculties, Kant argues that the moral formation of mankind – and thus the realisation of an ethical community – has to be supported by a ‘governmental plan’, and he describes war as the ‘greatest hindrance for morality’ (CF 7:92–93). Kant’s standard view thus entails three interconnected claims, i.e., the claim that a political community precedes peace, the claim that peace precedes an ethical community, and the claim that a political community precedes an ethical community.

In contrast to the aforementioned passages, where Kant argues – or at least suggests – that the realisation of peace in and through a political community anticipates the formation of an ethical community, Kant develops a second and, what we call, alternative view. In a sparing passage from the Religion he claims that the establishment of an ethical community based on laws of virtue guarantees perpetual peace: ‘Such is therefore the work of the good principle […] in erecting a power and a kingdom for itself within the human race, in the form of a community according to the laws of virtue that proclaims the victory over evil and, under its dominion, assures the world of an eternal peace’ (R 6:124). This alternative view also seems to surface in Perpetual Peace when Kant writes: ‘It can therefore be said, “Seek ye first the kingdom of pure practical reason and its justice [Gerechtigkeit], and your end (the blessing of perpetual peace) will come to you of itself”.’ (PP 8:378). The alternative view thus holds that an ethical community precedes peace.

Hence, the standard and the alternative view yield a dilemma concerning the historical-teleological sequence of establishing political and ethical communities with regard to peace: either the insurance of a political community has to be conceived as a condition for realising subsequently perpetual peace, moral reform and the establishment of an ethical community (I 8:21; I 8:26-28; CF 7:92–93; R 6:94–95), or the formation of an ethical community assures the realisation of perpetual peace (R 6:124; PP 8:378). The main burden of Kant’s views concerns not the question whether a political community anticipates an ethical community or vice versa, but the question whether peace precedes the establishment of an ethical community or vice versa. We therefore disagree with Rossi’s interpretation of the problem at stake:

‘In the […] Religion, [Kant] describes the ethical commonwealth as the community of virtue that assures perpetual peace. This suggests that a full establishment of the inner moral order of the ethical commonwealth precedes the creation of an external political order that will be effective for sustaining peace among nations. In […] The
Conflict of the Faculties, [...] Kant suggests the opposite relation: Securing an external political order of perpetual peace is a condition that needs to precede the establishment of an ethical commonwealth.’ (Rossi, 2005:§3.8)

By starting his interpretation with the alternative view and ending it with the standard view, Rossi generates the impression that the alternative view is to be taken as the point of departure and that the standard view is to be regarded as a deviation of the former. However, this is clearly not the case for, as we have shown, textual evidence mostly favours the standard view. Moreover, Rossi infers from the alternative view that an ethical community precedes a political community securing peace. However, Rossi does not give any further arguments for making this inference. Moreover, there seems to be no conclusive reason to make this inference: as far as we know, Kant never explicitly claimed that an ethical community precedes a political community. On the contrary, as we have seen, Kant repeatedly argued that without the foundation of a political community an ethical community cannot be brought into existence. What is important to keep in mind, is that the problem at stake concerns thus not the question whether a political community assures an ethical community or vice versa, as Rossi holds, but the question whether peace assures an ethical community or vice versa.

Before tackling our proposed solution for this dilemma, some final preliminary remarks should be made concerning the order between politics and morality in Kant’s philosophy – although it falls beyond the scope of this paper to deal with this in detail for ‘Even fundamental issues, such as Kant’s conception of the relation of political to moral philosophy, remain unclear’ (Wood, 2005:171). For Kant morality always precedes and has priority over politics in this sense that good politics always has to be rooted in morality. In Perpetual Peace Kant makes the distinction between the political moralist who treats the problem of constructing a peace-loving society as a technical problem that has to be solved by means of external and heteronomous legal rules, and the moral politician who approaches this problem as a moral problem appealing to the necessity of an inner and autonomous organisation of moral self-legislators (PP 8:370–380). There is no doubt that Kant prefers the moral politician: ‘I can indeed think of a moral politician, that is, one who takes the principles of political prudence in such a way that they can coexist with morals, but not of a political moralist, who frames a morals to suit the statesman’s advantage’ (PP 8:372). Accordingly, Kant further argues that morality has to precede true politics: ‘True politics can therefore not take a step without having already paid homage to morals, and although politics by itself is a difficult art, its union with morals is no art at all; for as soon as the two conflict with each other, morals cuts the knot that politics cannot untie’ (PP 8:380). As La Caze (2007:782) argues, morality and politics cannot conflict because morality ‘places limits on what can be done in politics’. Similarly, Castillo (2004:201) contends that only by taking up the perspective of the moral politician, can one avoid a conflict between morality and politics and can
one be assured that what has to be done from a moral point of view, will be realised in politics so that peace will be established on the basis of a harmony between morality and politics.

In Kant’s views on realising political and ethical communities with regard to perpetual peace, we can therefore discern a twofold interaction between morality and politics. First, morality functions as the most fundamental level from which true politics can spring off and grow. As we have just seen, morality functions as the ultimate decision-maker for politics – only the moral politician safeguards just politics leading to peace. Secondly, true politics can lift mankind’s inner morality to an even higher level: ‘it is not the case that a good state constitution is to be expected from inner morality; on the contrary, the good moral education of a people is to be expected from a good state constitution’ (PP 8:366). The main reason for this line of thought is that a good state constitution, viz. a republic and by extension a cosmopolitan league of republics, will refrain from waging war so that more government funding can be invested in education, instruction and culture, which function as tools for cultivating man’s moral habitus and hence man’s preparedness to enter into an ethical community. Furthermore, we concur with Anderson-Gold (2001:108) who defends Kant’s account of a political community as a necessary condition of an ethical community by arguing that ‘in order for ethical communities to fully promote universal ends, they must address the concerns of citizens of other states and so in this sense the goals of the ethical community presuppose an international legal order that at least minimally and provisionally expresses the interests of other communities.’ What is important to keep in mind, is that – without offering any clear-cut details – Kant interprets morality and politics as comprising a kind of reverberate circuit in this sense that both reinforce or presuppose one another.

Our interpretation of Kant’s historical-teleological views on establishing political and ethical communities with regard to peace gains support from the Vorarbeit zur Metaphysik der Sitten, where Kant suggests that only when freedom is secured externally by laws, inner moral freedom and virtuous dispositions will revive and will in turn enhance obedience to lawful coercion:

‘The solidly grounded peace by the greater intercourse of people amongst each other is that idea through which alone the transition of duties of right to duties of virtue is made possible, because when the laws secure freedom outwardly, the maxims can revive themselves also to govern inwardly according to laws, and conversely, these in turn relieve the influence to the lawful coercion through their dispositions, so that peaceful conduct under public laws and peace-loving dispositions [...], hence legality and morality find in the concept of peace the point of support of the transition from the Doctrine of Right to the Doctrine of Virtue.’ (VMS 23:353–354).\(^1\)

\(^1\)My translation. See also Kleingeld, 2006:493-494.
This kind of dynamic interaction between politics (rights) and morality (virtues) promotes the gradual transition from a legal – but fragile and provisional – peace in a political community based on self-interest triggered by the cunning of nature, to a moral and stable peace in an ethical community based on genuine moral interest triggered by pure practical reason. We believe that only by making a distinction between two concepts of peace, legal and moral peace, related to Kant’s political and ethical communities respectively, the burdens with Kant’s confusing historical-teleological views on establishing these communities with regard to peace can be resolved. To summarise the picture, a political community anticipates the establishment of a legal peace, which is only provisional since it is merely based on securing people’s external freedom and rights by means of legal coercion. If legal coercion drops out, people cannot rely on their own inner force to maintain peace since their moral dispositions are not yet fully developed. However, a provisional legal peace has a positive role to play for it stimulates education, instruction and culture, and as such prepares the formation of an ethical community. This in turn consolidates legal peace by upgrading it to a moral, lasting and hence truly perpetual peace: in an ethical community peace will also be based on securing people’s internal freedom and virtues in a non-coercive way.

In the passage quoted above, Kant speaks of a transition from the Doctrine of Right to the Doctrine of Virtue. It is a natural question to ask what kind of transition Kant has in mind and how this might relate to the above proposed interpretation. As is widely known, the Doctrine of Right concerns ‘the rational standards for externally coercive laws and the foundations of the human institution (called “civil society” or “the political state”) within which such laws have their place’, whereas the Doctrine of Virtue is ‘a theory about any human being’s regulation of its own conduct in accordance with self-given laws of reason’ (Wood, 2005:171). As such, the former is ‘concerned only with protecting the external freedom of individuals, and is indifferent to the incentives that lead them to follow its commands’, whereas latter is concerned with ‘the self-government of rational beings’ and hence also deals with ‘the ends people set and the incentives from which they act’ (Wood, 2005:143-144). Although both doctrines taken together constitute what Kant calls ‘morals’, they are to be seen as distinctive domains (of one science of morals) with their own distinctive grounds, viz. external, legal and enforceable versus inner, moral and non-enforceable freedom respectively. In the Vorarbeit zur Metaphysik der Sitten Kant therefore calls the transition from the Doctrine of Right to the Doctrine of Virtue a ‘transition from one order of things to another’ and argues that this is to be conceived as ‘a step [Schritt] in which each foot stands on its own particular – and from another distinguished – ground [Boden]’ (VMS 23:354).

Hence, the transition from the Doctrine of Right to the Doctrine of Virtue cannot be conceived analogously as the transition from the Groundwork of the Metaphysics of Morals and/or the Critique of Practical Reason to the

1 See also Rauscher, 2007:§1.
2 My translation.
Metaphysics of Morals, i.e., as a transition from the transcendental foundations of the use of pure practical reason to its field of application: the Doctrine of Right cannot be understood as the transcendental foundation for the Doctrine of Virtue since both are distinctive domains of the same field of application of pure practical reason, i.e., ‘morals’. Moreover, the transition from the Doctrine of Right to the Doctrine of Virtue via the idea of peace implies that, as La Caze (2007:794) puts it, ‘Kant did not think that virtue and right were necessarily co-implicated but instead had a hope that people would live according to the virtues of love and respect once right restrained politics’. Once again, this fits with our proposed interpretation of the establishment of political and ethical communities with regard to peace. Only when right restrains politics and the establishment of a political community assures a legal peace, i.e., a peace that coercively safeguards people’s external freedom and rights, can people fully develop their moral habitus and live up to the maxims of morality so that an ethical community is realised. This in turn consolidates and stabilises legal peace by turning it into a moral and truly durable peace because this peace will not only be based on mere self-interest in one’s external freedom and rights, but also on common moral dispositions concerned with one’s inner freedom and virtues.¹

Bibliography


Kant, I (1900-). Gesammelte Schriften. Berlin: Walter de Gruyter.


¹I would like to thank Bart Vandenabeele for his most generous support with regard to my research, as well as two anonymous referees for their invaluable comments. I also wish to thank the Scientific Research Foundation Flanders for their financial support.
20th Century
Foucault: The Cost of Truth and the Price of Transgression

Julio César Díaz

In several interviews and lectures he gave in the 1970’s, Foucault claimed that both his archeology of knowledge and his genealogy of power were akin to the critique of knowledge that Kant had undertaken. However, Foucault’s critique was not a critique of the conditions that make knowledge possible, of ‘what limits knowledge has to renounce transgressing.’ It was a ‘historical critique of the conditions in which knowledge is articulated by power.’

Power appears in multiple relations; therefore, a critique of power should focus on particular discourses, strategies, and practices (such as those of madness, and sexuality), rather than on a critique of formal structures with universal value. Foucault’s analyses called for the ‘eventualization’ (événémentialisation) and spatialization of power relations, by means of which ‘[specific] mechanisms of coercion and contents of knowledge [could be] identified.’ In his analyses, truth emerged both as being ‘produced by multiple forms of constraints’ and as ‘having coercive effects of power’ that configure systems of exclusions, prohibitions, divisions, and rejections.

Foucault saw the subjectification of human beings as an effect of truth. In an interview given to the German magazine Spuren (Foucault, 1996), Foucault pondered upon the following question: ‘[H]ow much does it cost the subject to be able to tell the truth about himself?’

This paper will address the following questions concerning Foucault’s notion of truth:

1. Since Foucault’s historical critique implied a spatialization of the relations between power and knowledge, how is truth affected by it?

---

2Foucault, Politics of Truth, 124.
3Ibid., 50.
2. If ‘truth is produced and transmitted under the control … of … political and economic apparatuses (university, army, writing, media),’¹ what could be to the cost of transgression, of ‘saying the truth’ without ‘being in the truth’?²

The Persistence of Truth

In the works of Michel Foucault there is an important series of reflections on truth and its history. According to Foucault, truth is not simply a disclosure or a correspondence between statements and facts.³ Truth is not the result of an analysis of costs and benefits, as the pragmatism suggests.⁴ Truth is not a language game detached from historical circumstances and alien to the strategies followed by its players.⁵ For Foucault, truth is the result of particular procedures and the effect of specific practices, of a power that configures a cognitive space and imposes itself upon the subjects, turning them into objects of knowledge.⁶ Sustained by multiple relations of power, truth is able to maintain spatial relations and to exercise specific effects of power upon subjects.⁷

In the interview *Truth and Power* (1980b), Foucault described truth in terms of specific processes of production and distribution. The following features characterize the political economy of truth in the Western culture: [1] Truth lies at the center of the scientific discourses produced by institutions; [2] Truth is dominated by political and economic powers; [3] Truth circulates as an object of diffusion and consumption; [4] Truth is produced and transmitted under a dominant, but not exclusive, control of few great political and economic apparatuses (such as the university, the army, writing, the media); and [5] Truth is at stake [l’enjeu] in every political debate and in every social struggle.

⁵‘Rules are empty in themselves, violent and unfinalized; they are impersonal and can be bent to any purpose. The successes of history belong to those who are capable of seizing these rules, to replace those who had used them, to disguise themselves so as to pervert them, invert their meaning, and redirect them against those who had initially imposed them….’ Foucault, M.(1977). ‘Nietzsche, Genealogy, Practice.’ In D. Bouchard (ed.). *Language, Counter-Memory, Practice*. Ithaca, NY: Cornell University Press, 151.
Clearly, for Foucault, truth is circularly bound to power systems that produce and sustain it.\(^1\) On the basis of this analysis of truth, Foucault suggested that the essential political problem for the intellectual is to ascertain ‘the possibility of a new politics of truth’ by changing the actual regime of truth. The task set for the intellectuals is not to detach truth from every system of power, but to detach the power of truth from the hegemonic forms in which it operates at the present time.\(^2\)

It was characteristic of Foucault’s analyses to conceive truth in terms of statements (énoncés),\(^3\) and to identify truth with a series of practices, techniques, and procedures by means of which truth is obtained and affirmed.\(^4\) In *Pouvoir et Savoir* (2001b), Foucault affirmed:

‘by truth I understand the entirety of procedures that allow at every moment... to utter [prononcer] statements [énoncés] that shall be considered true. There is no other supreme instance.’\(^5\)

This identification between true statements and the procedures that allow us to utter and sustain them entails significant consequences for Foucault’s reflections on truth and the history of truth. One of the most important consequences is the eventualization of truth. A statement can only be considered true with regard to a series of practices, within a specific field of knowledge [savoir], and within the realms of a particular conception of knowledge [épistémè]. Statements are ‘not ... incarnations of an essence but ... singularities.’\(^6\)

Foucault didn’t cast doubt on the existence of truth, as a skeptic could have done. What Foucault put in question was the persistence of truth, that is to say, the practices and mechanisms by means of which truth is maintained and empowered.\(^7\) Although Foucault has never used the terms existence and persiste in reference to truth, it is necessary to draw a distinction between the existence and the persistence of truth. While the existence of truth denotes a set of rules and principles that ‘enable [a person] to distinguish true and false statements,’\(^8\) the persistence of truth (per: intensive prefix + sistere: to remain, to stand firm) denotes an aspect of truth: its support, backing, or guarantee. To talk about the persistence of truth is to suggest that truth lacks its own power (pouvoir), that truth is powerless unless it is backed or empowered by a series

\(^1\)‘La vérité n’est pas hors pouvoir ni sans pouvoir.’ Foucault, Dits et Ecrits 2, 112.
\(^2\)Cf. Foucault, Power/ Knowledge, 131-133; Foucault, Dits et Ecrits 2, 112-114.
\(^3\)Foucault, Power/ Knowledge, 131-133; Archeology of Knowledge, 200-211; Dits et Ecrits 2, 873.
\(^4\)Cf. Foucault, Dits et Ecrits 2, 112.
\(^5\)Foucault, Politics of Truth, 55.
\(^6\)Specifically, Foucault raised questions on how “the true and the false are separated and specific effects of power [are] attached to the truth.” Cf. Power/Knowledge, 132.
\(^7\)Cf. Foucault, Archeology of Knowledge, 90-91; Dits et Ecrits 2, 875-6, 1544-45; Power/Knowledge, 131.
of practices and institutions. Without this support from power [le pouvoir],
truth is—in Foucault’s own statement— a disarmed truth, a ‘vérité désarmée.’
Foucault’s genealogies can be understood as analyses whose aim is to question
the persistence of truth, that is to say, to inquire into the conditions (the
historical apriori) that uphold, maintain, or support the truth. An important corollary of this distinction is that truth can no longer be
regarded as a-temporal. Once the practices and institutions configuring a
particular knowledge [épistémè] mutate, shift, or disappear ‘as the universal
foundation of all possible orders,’ the procedures that back up truth may
mutate accordingly. Truth has its own history, a history that reflects the
mutations and discontinuities in the épistémè of a particular culture.

I shall insist on that the distinction between existence and persistence is not
explicitly stated in Foucault’s works. I have drawn this distinction to point out
the epistemic problems showed by Foucault to those who aimed at writing a
‘universal’ history of truth. Foucault called our attention to these difficulties in
The Discourse on Language (1972b), where he made a distinction between statements that are within the realm of truth (“dans la vérité”) and those
statements that belong to the realm of truth and configure a particular practice (“sur la vérité.”) In a controversial passage from The Discourse on Language,
Foucault argued that Mendel ‘spoke the truth, but he was not within the truth
(dans la vérité) of contemporary biological discourse.’ According to Foucault,
to be in the realm of truth (‘dans la vérité’) means that a theory has been
furnished by the scientific discourse of its time. The expression ‘dans la vérité’
underscores how a theory derives its epistemic force (its capacity to classify,
order, and exclude) from such a support.

In conflict with the principles of the contemporary biological discourse,
Mendel’s theory lacked epistemic force [pouvoir]. In other words, Mendel’s
theory existed: it was conceived, developed, and tested. Yet it was not able to
persist because it lacked the support provided by the biological discourse and
institutions of its time. To understand the relationship between truth and the

---

1Foucault, Dits et Ecrits 2, 488.
2‘“Truth” is to be understood as a system of ordered procedures for the production, regulation,
distribution, circulation and operation of statements. “Truth” is linked in a circular relation
with systems of power which produce and sustain it, and to effects of power which it induces
and which extends it. A “regime” of truth.” Dits et Ecrits 2, 74, italics added; Peters, M.A.
Oxford Review of Education, 210-211.
Archeology of Knowledge, 219; Naissance de la Clinique, 1-8.
4Cf. Foucault, The Archeology of Knowledge, 219; Foucault, Naissance de la Clinique, 199-
200.
5Foucault, The Archeology of Knowledge, 224.
6A proposition must fulfill some onerous and complex conditions before it can be admitted
within a discipline; before it can be pronounced true or false it must, as Monsieur Canguilhem
might say, “within the true” Foucault, The Archeology of Knowledge, 224.
7‘It is always possible one could speak the truth in a void; one would only be in the true [dans
la vérité], however, if one obeyed the rules of some discursive ‘policy’ which would have to be
reactivated every time one spoke. Disciplines constitute a system of control in the production
practices that maintain the status of the truth requires us to examine the history of the scientific discourses and the power these discursive practices grant to particular theories.1 This is the task that Foucault carried out in his genealogies.2

The Spatialization of Truth

According to the distinction I have drawn in the previous section, the persistence of truth is the fundamental problem that Foucault set out in his works. Foucault often expressed this problem in spatial terms (e.g. territory, field, displacement, domain, region, horizon, limit, etc.), which organized the experience of knowledge.3 It is in this sense that we can talk about a spatialization of truth in Foucault’s writings. The spatialization of discourse in different epistemic fields allows us to see how truth depends not only on specific practices, but also on specific institutional frames. What is more, the spatialization of truth shows the limits within certain practices and therefore the limits of the effects of truth.4

For instance, in the field of clinical medicine, Foucault noted that in the 17th century, the ‘natural locus’ of a disease was the family.5 As a consequence of a series of shifts in the Western configuration of knowledge, at the beginning of the 18th century, the hospital appeared as a privileged space where diseases found their ‘natural locus.’6

To reflect upon truth and its effects in terms of a spatial configuration brings some significant epistemic consequences.7 It leads us to [1] distinguish the locus of a true discourse organized in different disciplines; [2] define the legal domain that allows particular practices (e.g. the series of rules and regulations that allowed the hospital to become the privileged space of clinical regard); [3] of discourse, fixing its limits through the action of an identity taking the form of a permanent reactivation of the rules.’ Ibid., 224.

1Cf. Foucault, Power/Knowledge, 96.
5‘Le lieu naturel de la maladie, c’est le lieu naturel de la vie –la famille: douceur des soins spontanés, témoignage de l’attachement, désir commun de la guérison, tout entret en complicité pour aider la nature qui lutte contre le mal, et laisser le mal lui-même se déployer dans sa vérité: le médecin d’hôpital ne voit que des maladies louches, altérées, toute une tératologie du pathologique…’ Foucault, Naissance de la clinique, 16.
6Cf. Foucault, Naissance de la clinique, 39.
7‘People have often reproached me for these spatial obsessions, which have indeed been obsessions for me. But I think through them I did come to what I had basically been looking for: the relations that are possible between power and knowledge. Once knowledge can be analysed in terms of region, domain, implantation, displacement, transposition, one is able to capture the process by which knowledge functions as a form of power and disseminates the effects of power.’ Foucault, Power/Knowledge, 69.
specify the field in which certain statements are validated; and [4] reflect upon the limits and effects of truth before our transgression of those limits.1 According to Foucault, a history of truth must necessarily address all these aspects.2

It has been suggested that Foucault’s spatial analysis of truth is a form of relativism, for it assumes that truth is not universal.3 The idea that truth is universal suggests that truth is valid at all times and in all places. Throughout its history, Western philosophy has attempted to set the conditions under which truth appears as universal. This project has been vehemently attacked by Foucault.4

The claim of the universality of truth loses its ground when it is confronted with the demand of specifying the particular practices from which truth derives. For Foucault, the definition of truth necessarily implies the specification of the practices from which truth acquires its status. The difficulties Foucault pointed out remain unsolved: Can we talk about truth (e.g. the validity of true statements or énoncés) beyond the discursive practices that support it? Under what circumstances can we ignore the procedures from which truth derives its validity and force (effets réglés du pouvoir)? What practice could justify the writing of such a history of truth?

Dialectics, Fiction, Transgression

To emphasize the spatial character of truth (its locus, its limits) is to acknowledge that the universal character of truth is affected by such a spatialization. This concession may not carry the negative consequences that Foucault has often been charged with: relativism, skepticism, an inability to justify ethical choices, and a bias judgment of different political systems as equally repressive regimes.5 The spatialization of truth allowed Foucault to identify limits, points of convergence, and particular reconfigurations of...

---

1 In Truth and Power (1980b), Foucault distinguished different aspects of a regime of truth: ‘the type of discourse which it accepts and makes function as true; the mechanisms and instances which enable one to distinguish true and false statements, the means by which each is sanctioned; the techniques and procedures accorded value in the acquisition of truth; the status of those who are charged with saying what counts as true.’ Ibid., 131.


4 Cf. Foucault, Dits et Ecrits 2, 1258-1260, 1408-1412; Foucault, Naissance de la Clinique, 56.

knowledge [savoir].¹ The distinction we have drawn between the existence and the persistence of truth may be relevant here as well. Although truth may exist at all times and in all spaces, its effects may not equally persist beyond a particular historical time. In order to persist, truth must be endowed with a force (puissance) that allows it to produce effects of power over the subjects. Without the power relations [relations du pouvoir] that sustain it, truth may be completely powerless, ‘une vérité désarmée.’²

Yet, the spatialization of the relation between truth and power raises some significant epistemic problems. As we have seen in the previous section, Foucault contends that political and economic powers control the production of truth, its place, circulation and distribution. We can ask what the price of transgressing these controls could be, what the price of ‘saying the truth’ without being ‘in the truth’ (dans la vérité) is. In an interview given to the German magazine Spuren,³ Foucault himself raised the following questions:

‘While the historian of science in France busied himself primarily with the problem of the constitution of scientific objects, I asked myself another question. How does it happen that the human subject makes himself into an object of possible knowledge, through which forms of rationality, through which historical necessities, and at what price? My question is this: how much it cost the subject to be able to tell the truth about himself? How much it cost the subject as madman to be able to tell the truth about himself?’⁴

Positing the problem of truth in terms of the specific procedures that sustain and empower truth, Foucault had in mind the possibility of desubjectification and transgression.⁵ Foucault has repeatedly insisted on the possibility and need of intellectual transgression. By intellectual transgression, Foucault meant the experience of desubjectification through the analysis of the constitution of the subject (that is to say, through the study of specific techniques of control, surveillance, and identification of the subject).⁶ Foucault described the disclosure of the ‘lowly origins’ of truth as inevitably linked to the experience of liberating the subjects from the hegemonic powers that have constituted them. Thus, the task of the intellectual entails a particular kind of transgression:

¹Cf. Foucault, Naissance de la Clinique, 77-78.
³Cf. Foucault, Foucault Live, 348-362; Power/ Knowledge, 131; The Archeology of Knowledge, 219-220.
to rethink the history of truth not simply as ‘the reward of free spirits, the child of protracted solitude, [or] the privilege of those who have succeeded in liberating themselves,’ but rather as the history of specific procedures, mechanisms of exclusion, and means of subjectification of individuals. In other words, it is the task of thinking truth as ‘…a thing of this world: [truth] is produced only by virtue of multiple forms of constraint. And it induces regular effects of power.’

Foucault's studies of these historical conditions for the production of truth had an ultimate goal, which is ‘[to detach] the power of truth from the forms of hegemony, social, economic and cultural, within which it operates at the present time.’ If an intellectual transgression is possible, its price may be:

1. To conceive and to adopt the non-dialectical language of the limit, that is to say, to challenge what is said by those who are in a position of authority and empowered to speak. Foucault described this search as a quest for a ‘language stripped of dialectics,’ a language that ‘flee[s] from the horror of reducing being to totality’ and allows us to reflect upon ‘our relation to truth.’

2. To give up the status of ‘bearer of universality’ coming from the time when the intellectual ‘was acknowledged the right of speaking in the capacity of truth and justice.’ In carrying out his/her task, the intellectual does not deny truth, but rather insidiously questions the source of its effects of power. The intellectual does not deny the universal character of truth; instead, he/she raises questions about the status of such universality.

3. To affirm the ‘fictional’ nature of one’s writings. For Foucault, to write ‘fictions’ was a way of inducing subtle effects of truth that might reconfigure the fields in which truth was placed at his time.

---

1 Foucault, *Power/Knowledge*, 131.
4 Foucault, *Power/Knowledge*, 133.
5 Foucault, *Language, Counter-memory, Practice*, 44.
6 Ibid., 41.
7 Ibid., 43.
9 Ibid., *Power/Knowledge*, 126.
11 I am well aware that I have never written anything but fictions. I do not mean, however, that truth is therefore absent. It seems to me that the possibility exists for fiction to function in truth, and for bringing it about that a true discourse engenders or ‘manufactures’ something that does not as yet exist, that is, ‘fictions’ it. One ‘fictions’ history on the basis of a political reality that makes it true, one ‘fictions’ a politics not yet in existence on the basis of a historical truth.” Foucault, *Power/Knowledge*, 193.
In that sense we have to understand Foucault’s statement, ‘je me rends bien compte que je n’ai jamais écrit que des fictions.’ His ‘fictions’ were not false or ‘beyond truth,’ ‘hors vérité.’ His ‘fictions’ were placed within the realm of truth (‘dans la vérité’), from where they were able to ‘induce an effect of truth’ [induire effets de vérité] and to ‘stir up, produce a political reality that rendered it real.’

Foucault’s writings on the prison, on madness, and on the clinical medicine were ‘fictions’ that aimed at creating some sort of ‘interference’ against the hegemonic powers controlling truth.

Concluding Remarks

For Foucault, to write the history of truth meant to place truth in a particular realm, to fix ‘le réseau de son space et de son développement.’ This historical critique would make explicit the procedures and ‘the conditions in which knowledge is articulated by power.’

Foucault’s spatialization of the relation between power and truth posited a central problem: the persistence of truth. From the perspective of the persistence of truth, the task of the intellectual requires a particular kind of transgression. The task of the intellectual is no longer merely to establish the universal conditions of the existence of truth, but to determine the historical conditions and realms in which truth is sustained by power [pouvoir]. The spatialization of truth and its effects make explicit the limits of these effects. For Foucault, the price of intellectual transgression is to break away from the a-temporal notion of truth and the discourses that characterizes the subject as universal.

Thus, in Foucault’s philosophy, the quest for truth became a series of reflections upon particular strategies and techniques of surveillance and control in which truth showed its persistence, its effects upon the subject, and ‘its vocation to exclusion.’

---

1. ‘Suscite[r], fabrique[r], …une réalité politique qui la rend vrai.’ Foucault, Dits et Ecrits 2, 236.
2. Foucault, Naissance de la Clinique, 155.
4. ‘Metaphorising the transformations of discourse in a vocabulary of time necessarily leads to the utilization of the model of individual consciousness with its intrinsic temporality. Endeavoring on the other hand to decipher discourse through the use of spatial, strategic metaphors enables one to grasp precisely the points at which discourses are transformed in, through and on the basis of relations of power.’ Foucault, Power/Knowledge, 69-70; Cf. Foucault, Politics of Truth, 124-125; Foucault, Naissance de la Clinique, 175; Cf. Keenan, T. Political Theory, 8-9.
Clearly, in Foucault’s transgressive reflections, the persistence of truth was at the center of the interrogation. It is from this center that we should reassess truth and its history.

**Bibliography**


Norris, C. (1994). ‘What is Enlightenment?: Kant According to Foucault.’ In G.
CHAPTER THIRTEEN

Craig and His Concept of Eternity: 
A Critique from the Standpoint of the Kālām

Engin Erdem

The issue of divine eternity which was dominant in medieval philosophy and theology also has a very important place in the debates of modern philosophy of religion. Looking at the views of the medieval philosophers and theologians such as Ibn Sina (d. 1037), St Anselm (d. 1109), and St Thomas Aquinas (d. 1274), we see that they interpret divine eternity as timelessness, being outside of time. According to these philosophers there is a sharp distinction between God’s mode of being and that of the creatures. A created being has its life in a succession; past, present and future are real parts of its life. However, because of the fact that God is a perfect, simple, necessary and an immutable being, there is no change and succession in His life. So Boethius (d. 524) defines eternity, as having an illimitable life all at once. (1981, 430) Here the term ‘all at once’ means not a moment of time but the absence of temporal succession. (Helm, 2007) Also, Anselm depicts God’s relation to time as follows:

Thou wast not, then, yesterday, nor wilt thou be to-morrow; but yesterday and to-day and to-morrow thou art; or, rather, neither yesterday nor to-day nor to-morrow thou art; but simply, thou art, outside all time. For yesterday and to-day and to-morrow have no existence, except in time; but thou, although nothing exist without thee, nevertheless dost not exist in space or time, but all things exist in thee. For nothing contains thee, but thou containest all. (1962, 71)

For the medieval philosophers, whatever is in time is bounded by it; it cannot stop the process of change and of time; so, it is subject to time not its master. (Helm, 2007) Therefore, God must be timeless.

As for the modern philosophers of religion, their approach is radically different from the medieval philosophers. These philosophers oppose the classical view of eternity, saying that a timeless being cannot be God of religion, especially ‘the Christian God’. (Davies, 1983, 11) According to many of them, God, in scripture, is described as a loving, suffering, redeeming being, in brief, an acting being. But if God is unchangeable and immutable he cannot do such activities; thus, a being that is not changeable cannot be an agent. Hence, it is necessary to ascribe change and temporal properties to God and
God must be temporal. What does God’s being temporal mean? Almost all of modern philosophers of religion answer this question in different ways. For example, Swinburne, one of the leading defender of divine temporality, holds that eternity means to have an everlasting life in time. He says that God exists throughout all periods of time; God exists now, he has existed at each period of past time and he will exist at each period of future time. (1994, 137) God exists in time but there is no beginning or end for Him; so ‘a being who is both backwardly and forwardly eternal we may term an eternal being.’ (1993, 218) Similarly, Wolterstorff maintains that God exits in time and he does have a history like the creature. (2001, 211) Although modern philosophers of religion have relatively different views on the nature of divine temporality, they all share the same opinion: God cannot be a timeless being.

As it is seen, there are mainly two views about the nature of divine eterni-
ty: divine timelessness and divine temporality. But we see that there is also a third way which is held by William Lane Craig. Craig has a distinguished place among modern religious philosophers. According to him, God is neither timeless nor temporal; God’s life has two stages: the first is timeless and the second is temporal. It seems that, as we shall see below, the turning point in his view of eternity is creation. For he claims that God is timeless without creation and than becomes temporal with creation. (2000, 152; 2001, 236) Why is God timeless; more importantly why did he become temporal when he created the universe? In what follows I shall deal with these questions in turn.

Let me begin with the first question. I think the answer of this question is closely connected with Craig’s concept of creation ex nihilo; because, according to him a robust/strong doctrine of creation implies that God is the Creator of everything except Himself and the universe was created a finite time ago in the past. (2004, 161) To prove such an idea of beginning for the universe he utilizes many scientific and metaphysical arguments. Especially the cosmological argument plays a key role in his thought. According to him, some Muslim thinkers such as al-Kindi and al-Ghazali introduced a different version of the cosmological argument, which he calls ‘The Kalām Cosmological Argument’ (2000). According to this argument, God is not only ontologically prior to the universe but also He temporally precedes it. The argument can be formulated as follows:

I. Everything that begins to exist has a cause of its existence.
II. The universe began to exist.
III. Therefore, the universe has a cause of its existence. (2000, 48-49)

For Craig, the crucial premise in this argument is the second one. He asserts that the impossibility of actual infinity, The Big Bang Theory, Thermodynamics and other many scientific evidences show that the universe has a beginning. (2004, 219-248) Introducing those arguments in detail he tackles the question whether the beginning of the universe also entails the beginning of time. According to him, on the relational theory of time the universe was created not in time but with time. (1979) The view that there was
an empty time before the beginning of the universe gives rise to an old metaphysical question: ‘Why did not God create the world sooner?’ (2001, 21) Craig argues that this question is unanswerable in the absolute theory of time. (2001, 31) So there is no time ‘before’ the beginning of the universe; the first event, that is, the creation of the universe also indicates to the first moment of time. (2001, 31; 1979) ‘When the first event occurred, the first moment of time began.’ (1979) Craig concludes that time did start with the beginning of the universe and without creation there was no time but God only; therefore, in the first stage, that is without creation God is timeless. (2001, 31)

Now, we are coming to the second question, ‘Why God becomes temporal with creation?’ According to Craig, as I said above, God is timeless without creation and if God did not will to create the world he would exist timelessly. (2000, 152) In other words, God could have never stood in temporal relations with a temporal world had he not willed to create the world. But Craig says that God has willed from eternity to create and to become temporal at that moment. (2000, 152) Because of the fact that with creation God enters into a new real relation which he did no have before He changes and becomes temporal. (2000, 152) Craig summarizes his view as follows:

1. God is creatively active in the temporal world.
2. If God is creatively active in the temporal world, God is really related to the temporal world.
3. If God is really related to the temporal world, God is temporal.
4. Therefore, God is temporal. (2001, 141)

Thus, for Craig, ‘the first event is the event of creation, the moment at which the temporal phase of God’s life begins.’ (2001, 31) I think what has been said so far outlines sufficiently Craig’s main thesis of eternity. Now, in light of the discussions between the classical and modern philosophers on the nature of divine eternity, we can say that Craig holds a hybrid view of eternity, partly timeless and partly temporal. It is a well known fact that, taken separately, each of the classical and modern interpretations of eternity brings about different kinds of problems, either philosophical or theological. But it seems that a mixed type of eternity held by Craig gives rise to further difficulties.

Firstly, I think the main problem arising from a hybrid view of eternity is the question that how the relationship between the timeless and temporal parts of divine eternity can be explained. As Leftow said, the expression that ‘God becomes temporal’ (1997, 259) means that God is timeless first and later becomes temporal. So God’s timeless phase comes earlier than His temporal phase, but the term ‘earlier’ or ‘before’ shows that there was a time before God’s becoming temporal and He did exist in that time. (1997, 259) And this means that God had been already temporal before having become temporal. Another problem is related to the state in which God decided to become temporal. If God, as Craig argued, did decide to become temporal, it might be asked: when did he do so. ‘He could not’, says Leftow, ‘have done so
timelessly, for then He would have had to become temporal... If He did so at any time it was then too late. As He was already at that time, He was already temporal.’ (1997, 259-260) It is seen that Craig’s idea that God decided to become temporal requires a time in which God already has been.

The second problem I would like to discuss concerning Craig’s view of eternity is the question that whether the act of creation really requires God’s becoming temporal. I think it might be helpful on this issue to look at the views of the defenders of divine timeless eternity, especially to whom Craig refers in order to support his concept of creation ex nihilo, such as al-Kindi (d. 873) and al-Ghazali (d. 1111). We see that those thinkers never approve of the view that God changes and becomes temporal with creation. To these philosophers, God, as a necessary, simple, and an immutable being by definition is timeless and there cannot be any change in Him, before or after creation. Let us take al-Ghazali as an example. He criticizes the Muslim Aristotelian philosophers, al-Farabi (d. 950) and Ibn Sina, from a religious point of view and objects to their view of eternal creation which implies that God is only ontologically prior to the universe. (1997, 31) He maintains that without creation there was only God and than the universe began to exist in accordance with God’s timeless will. (1997, 31) But he also argues that the inception of the universe does not require God’s becoming temporal. (1962, 104) For, according to him, only a timeless being can create temporal beings and can have an absolute control over them.

In order to better understand his approach to this issue I would like to touch upon his views on the relationship between God and time. He defines time, like Aristotle, as the measure of motion in terms of before and after. (1990, 172) In view of the fact that he sees an essential relation between time and motion, a question as to divine eternity might be, basically, answered in accordance with the answer given to the question of whether God has been subjected to a change or not. (Erdem, 2006, 3) Al-Ghazali tries to give an answer to this question by pointing out the radical difference between God’s mode of being and those of the created beings. (2006, 3) For him all the creatures are temporally created (hadith), which means that they came into existence after they had not existed previously. (1962, 25; 1997, 61) The coming into existence of the temporally created is either possible or impossible; its being impossible is not possible, because what is impossible cannot come into existence. (1962, 25) Hence it is necessary for a temporally created to be possible before it exists; a possible being is at the same of level with regard to existence and non-existence. (1962, 25) So, it needs a cause/decider that would prefer its existence to its non-existence. (1962, 25-26) But, since the chain of causes cannot continue infinitely, there must be a first cause which does not require any other cause but itself. (1962, 35) Thus, this first cause, with regard to its mode of being does not need to any other cause, is said to be a necessary being. A necessary being is different from a temporally created being, not only because it does not require any cause except itself but also it is not subjected to any motion or change. (2006, 3) Because, for him, motion is the actualization of a potentiality; in order for a motion take place an agent is required to actualize the potentiality. (1990, 172) So if a change and a motion were to
Craig and His Concept of Eternity: A Critique from the Standpoint of the Kālām

happen in God, this would mean that God had a potentiality and thus God would need an agent apart from Himself to put this potentiality into actuality. (2006, 3-4) In another words, if God changes, He would be at the same level of being with the created and He cannot be regarded as a First Cause. However, according to al-Ghazali since God is a necessary being, there cannot be any potentiality and any change in Him and therefore He must be timeless. (2006, 4) He expresses his view thus:

God, just like He was in eternity, in the infinite pasts, is always the same today. Just like He was before He created the universe and heavens, He would be exactly the same in the infinite future, too. Because change and alteration can not be ascribed neither to His being nor to His attributes. If one of His attributes were to change, He would be imperfect or flawed; therefore, be imperfect and would be in need of perfection and excellence. He who is in need can not be God. (1969, 112)

For al-Ghazali, God’s essence and all of His attributes, as well as His will, are timelessly eternal. (1962, 142) God decides timelessly to create the world in eternity. (1962, 104) God’s contemplation of unactualized possibilities and His decision to actualize one of them is one timeless act. ‘This does not mean that different phases of that act cannot be distinguished, but such a distinction can only be a conceptual, not a temporal distinction. Consideration of possibilities is logically prior to actualizing one of them, but both contemplation and actualization are one eternal act of the divine nature, if God is timelessly eternal.’ (Helm, 1988, 179) To al-Ghazali, even though God’s decision to create is a timeless act, the effects of this decision come into being temporally. But the temporality of the effects does not require the temporality of the cause. For, God’s act of creation is a different kind of causation, it is a timeless causation. Timeless causation is a relation of a cause and an effect between the Creator and the creation. (Markus, 2004, 32-33) To understand timeless causation, I think it might be helpful to compare, as Davies does, the relation between the cause and the effect with the relationship between a teacher and a student. When the student learns some truth he changes from a state of ignorance to state of having knowledge. It is a real change for her/him. But the learning of the student does not require a change in the teacher. Similarly, it might be coherently thought that God’s creation does not necessarily imply a change in Him but only a change in the creature. (Davies, 1993, 147)

At this point it might be asked that why al-Ghazali and other medieval philosophers insist so much on God’s immutability and timelessness and why they do not accept the view of God’s becoming temporal. To these philosophers, the attributes of necessity and timelessness are the marks of the Creator and the contingency and the temporality are the signs of the creature. For that reason, any explanation which implies God’s being temporal, in fact, is not a philosophically coherent explanation. Because in such a case, it would
be nearly impossible to make any distinction between God and His creature and this would lead to an anthropomorphic conception of God. So from al-Ghazalian point of view, Craig’s idea that God created the universe and became temporal with creation, in fact, means that God created the universe at the expense of His perfection. Craig might reply to this, saying that God ‘willingly’ did decide to become temporal at the moment of creation. (2000, 152) But such a response also seems questionable. Because questions such as follows always reasonably be asked: ‘Is it really possible that a being can change his mode of being by her/his own will?’ and more importantly, ‘For what reason did a perfect and a timeless being abandon His mode of being?’ I think it is very difficult to find any philosophically coherent answer to these questions in Craig’s view of eternity.

Lastly, I would like to add one point. Metaphysically thinking, an effect is ontologically dependent on its cause, not vice versa. In the classical view of eternity, God is conceived as a First Cause and everything apart from Himself wholly dependent on Him. Accordingly, the classical philosophers argue that the act of creation does not imply a real change in God but only in the creature. But in Craig’s conception of eternity, the effect is so much powerful that it can change the mode of being of its cause substantially. I think such a view brings about the question: ‘Which one, in fact, is the real cause, God or the universe?’

In conclusion, in the classical philosophy and theology God is conceived as an ultimate principle. In this view, God is the First Cause and everything except him came into being thanks to His act of creation. The fact that God is the First Cause means that there is a radical, sharp distinction between God and his creatures in terms of mode of being. For that reason, medieval philosophers never approve of the idea of God’s becoming temporal and try to explain God’s relationship with the temporal world in accordance with this metaphysical framework. But in Craig’s view of eternity, because he concedes that God becomes temporal with creation, it is very difficult to explain, at least after creation, the difference between the Creator and the creature. Likewise, his idea that God became temporal because of His relation with temporal beings is controversial. For, if God, as Craig claims, becomes temporal due to His relations with temporal beings it might also be thought that He becomes spatial due to His relations with spatial things. Of course, God might be conceived as a temporal and even a spatial being but it is clear that such a conception of God is more anthropomorphic than the timeless one.

Bibliography

Craig and His Concept of Eternity: A Critique from the Standpoint of the Kālām


CHAPTER FOURTEEN

Remarks on the Poetic Quality of Wittgenstein’s Philosophy

Chrysoula Gitsoulis

Wittgenstein’s intense preoccupation with language, his recognition of the importance of investigating the various ways in which we put linguistic expressions to use, expressions seemingly unimportant and trivial, is, arguably, what sets his philosophical works apart from so many other works that came before or after - which is why, more than any other philosopher, Wittgenstein is responsible for the celebrated ‘linguistic turn’ in philosophy. Meticulous and painstaking investigations into the ‘workings of language’ revealed, for Wittgenstein, the extent to which ordinary language is bound up in philosophical problems. A striking aspect of these investigations is the sense of caginess, even eeriness and spookiness, they arouse in the reader. In this paper, I offer four reasons to account for this feature of his writing, reasons that will make it apparent that the oracular, poetic quality they give rise to is not merely accidentally, as some have thought, but necessarily tied to his philosophical method.

Pseudo-problems

The sense of caginess (and in turn eeriness) that is generated by Wittgenstein’s style of writing can be traced in part to what he took to be a crucial function of his philosophical method: to dissolve (not solve) what he deemed to be pseudo-philosophical problems1; to show why they are illegitimate. Pseudo-problems, Wittgenstein felt, originate from an impoverished diet of examples of how words/sentences are used.

A main cause of philosophical disease – a one-sided diet: one ourishes one’s thinking with only one kind of example. [PI §593]

So, to deconstruct the pseudo-problems, Wittgenstein felt it necessary to present a rich battery of examples that remind the philosopher of the various uses of words/sentences. Only by bringing into view trivial and well known facts about our use of language, Wittgenstein felt, can we untie the knots in our understanding that give rise to pseudo-philosophical problems. Wittgenstein brings these facts into view by inventing language-games, comparing them, using diagrams and drawings, appealing to metaphors and jokes, etc., without, however, or rarely, explaining what the point of these examples is, or how they relate to each other. This style of writing creates a feeling of caginess because conventional methods of guiding the reader (crisp arguments, linear structure, etc.) are aborted; examples are simply placed before us, ‘criss-crossing in every direction’. [PI, Preface] We have the impression of someone on the prowl, who has lost his way. We want to stop him and ask: What are you looking for? But Wittgenstein hardly ever reveals where he is heading. We need to piece things together for ourselves, like the assemblage of a jigsaw puzzle, where no instructions have been on how to proceed.

Demystification

The caginess is also due to the fact that Wittgenstein felt that drawing out morals for the reader was not necessary for achieving his desired effect: demystification through the deconstruction of pseudo-problems. Wittgenstein allows the truth of what makes his philosophical method effective speak for itself. Just as a work of literature (a novel) can produce moral effects not by preaching, but simply by being, so too, Wittgenstein felt, that he could achieve his aim – disintegration of pseudo-problems – by presenting a rich battery of examples, vividly described, of language-in-action, without having to explicitly draw any morals for us.

If only you do not try to utter what is unutterable then nothing gets lost. But the unutterable will be – unutterably – contained in what has been uttered. ¹

The moral potency of literature attests to this. In order for a work of literature to have moral effects, it is not necessary that it present a theory of morality.² Nor is it necessary that it portray morally idealized subjects following rigid rules. Indeed, its moral potency is likely to be greater when it portrays morally mixed characters, much like the average viewer (as you find in ancient tragedies), and vividly described crises and struggles through which they pass, leaving us to draw the moral for ourselves. As Aristotle masterfully recognized in the Poetics, the idealized subject, free of common faults, loses his/her ability to engage our attention, and in turn our sympathy; the one who is like us, who we can relate to emotionally, and hence sympathize with, is the

¹In a letter from Wittgenstein to Paul Engelmann dated 9/4/1917. See Engelmann (1967).
²This point has been defended at length by Martha Nussbaum in (1986) and (1990).
best agent for imparting a moral message. So too, I believe, Wittgenstein recognized that the roughness and inexactness of commonplace speech, with all its imperfections, set before us in numerous examples, and not a neat philosophical theory, still more one couched in an idealized language, could serve as the best agent for producing the therapeutic (and arguably moral) effects he was after. And, for those effects to be produced, he felt, no morals need be explicitly stated. This contributes to the oracular, poetic quality of his work, and of course, once again, to the sense of caginess.

Transformation

The sense of caginess that is generated in Wittgenstein’s works can also be traced to the following crucial function of his philosophical method: to transform the nature of the reader, to force him into a kind of conversion. This conversion crucially involved unlearning certain bad habits. This is why he says: ‘Working in philosophy … is really more a working on oneself. On one’s own interpretation. On one’s own way of seeing things.’ [CV p. 16] And why, in the Preface to the Philosophical Investigations, he says ‘I should not like my writing to spare other people the trouble of thinking, but to stimulate someone to thoughts of his own’. Wittgenstein presents a rich battery of examples whose purpose is to untie the knots in our understanding that give rise to pseudo-problems. But the reader must wrestle with these examples, he must work certain things out for himself, to see the moral that stands behind them: how misinterpretations of language have led him astray. Active participation from his reader, Wittgenstein felt, was necessary for his work is to have a therapeutic effect on him, to uproot bad habits.

The bad habits in question can be traced, at root, to an unbridled urge to generalize. Allowing this urge free rein leads to the sort of intellectual dishonesty involved in generating pseudo-problems and pseudo-theories. For it is mainly this that generates pseudo-problems: a failure to keep the various senses and functions of words/sentences apart in our minds. Mental sloth/laziness drives the philosopher to assimilate what should be kept apart (‘as if the sense were an atmosphere accompanying the word, which it carried with it into every kind of application’ [PI #117]). To combat these bad habits, we must remain in constant vigilance of the seductive and overwhelming powers of language; we must be honest with our selves when we are in the grip of a conceptual tension (‘This isn’t how it is … But this is how it has to be!’), and look carefully into the ‘workings of language’ [PI §109], to get to the source of it. We must ‘distrust language’ [NB] in the sense of guarding from an instinctive or habitual urge to generalize, and reflect more carefully on the workings of language when we are confused.

Of course, this is not easy work:

‘Philosophy is a battle against the bewitchment of our intelligence by means of language’. [PI §103]
The philosopher is ‘engaged in a struggle with language’. [CV p. 12]

The struggle involves not so much a battle of the intellect, as of the will: the will to resist an unguarded urge to generalize (more specifically, ‘urge to misunderstand’ [§109]), one of the most deep-seated human urges.

What makes a subject hard to understand – if it’s something significant and important – is not that before you can understand it you need to be specially trained in abstruse matters, but the contrast between understanding the subject and what most people want to see. Because of this the very things which are most obvious may become the hardest of all to understand. What has to be overcome is a difficulty having to do with the will, rather than with the intellect. [CV p. 17]

As I have often said, philosophy does not call on me for any sacrifice, because I am not denying myself the saying of anything but simply giving up certain combination of words as senseless. But in another sense philosophy demands a renunciation, but a renunciation of feeling, not of understanding. Perhaps that is what makes it so hard for so many people. It can be as hard to refrain from using an expression as it is to hold back tears or hold in anger. [MS 213, 406]

What we need to do, then, is strengthen the will to resist the urge to misunderstand. This, as Aristotle recognized to be the case for any good habit, is not something that can be acquired through teaching, but only through practice.

Developing good habits, of course, is not easy work, for the grip of a false interpretation is very difficult to shake off once that interpretation has become sedimented in our ‘forms of life’: to the extent that it has become sedimented, it becomes more difficult for us to challenge it.

Getting hold of the difficulty deep down is what is hard. Because if it is grasped near the surface it simply remains the difficulty it was. It has to be pulled out by the roots; and that involves our beginning to think about these things in a new way. The change is as decisive as, for example, that from the alchemical to the chemical way of thinking. The new way of thinking is what is hard to establish. [CV p. 48]

Wittgenstein would have agreed full-heartedly with George Orwell when he wrote: ‘An effect can become a cause, reinforcing the original cause and producing the same effect in an intensified form, and so on indefinitely. Language becomes ugly and inaccurate because our thoughts are foolish, but the slovenliness of our language makes it easier for us to have foolish
thoughts.' In this way, a mythology becomes ‘embedded in our language’ [MS 213, 434].

Moreover, we drill ourselves so often and so make a habit of interpreting words or sentence-forms in terms of standardly assigned meanings/functions that we naturally stretch these interpretations into all contexts where these expressions or sentential-forms are found, and until something compels us to reconsider our interpretation, we cease to think twice about it. In this state of contentment, we fail to notice, and so take steps to free ourselves from, the forces of language that hold us captive; our complacency conceals our shortness of vision.

A picture held us captive. And we could not get outside it, for it lay in our language and language seemed to repeat it to us inexorably. [PI §115]

But notice that Wittgenstein says we ‘could not get outside it’, not ‘cannot get outside it’. And indeed, the point (as the past tense ‘could not’ suggests) we must not forget is: the processes that have lead to our bewitchment are reversible. Though philosophical writing has its share of bad habits that spread by imitation and tradition, these habits can be reversed, by unraveling the linguistic confusions that led to them.

… philosophy unties the knots in our thinking, which we have tangled up in an absurd way; but in order to do that it must make movements which are just as complicated as the knots. [PR Part 1 #2]

By tracing the usage of the concepts that we have ‘tangled up in an absurd way’, it becomes possible to see how they became entangled and thus to disentangle them.

Language-games

Finally, the sense of caginess or eeriness that is generated in Wittgenstein’s work is tied, more specifically, with the way in which his philosophical method operates (and not merely, as I tried to show in 1-3, with the aim or goal of his method); in particular, it is due to the construction of hypothetical language-games. Hypothetical language-games involve imaginary uses of language that are meant to be compared to actual language-games. They involve constructing ‘objects of comparison which are meant to throw light on the facts of our

---


language by way of similarities and dissimilarities’. [PI §130] A hypothetical language-game is a method of instruction, and an indispensable one at that:

Nothing is more important for teaching us to understand the concepts we have than constructing fictitious ones. [CV p. 74]

What is involved in constructing a ‘fictitious concept’ is simply assigning a new role to a concept (one distinct from its actual role), rearranging the phenomena of language, so to speak, and then exploiting the dissimilarity between the invented and actual language-games to bring out a new angle on the concept. Hypothetical language games involve abstracting concepts from their normal circumstances of application. ‘It disperses the fog to study the phenomena of language in such imagined uses of language’, Wittgenstein remarks, ‘for here we can command a clear view of the aim and functioning of words’. [PI §5] Assembling reminders, rearranging facts, and assimilating pictures 'alter our way of looking at things' [PI §144]; they help put things into view.

One of the most important methods I use is to imagine a historical development for our ideas different from what actually occurred. If we do this we see the problem from a completely new angle. [CV p. 37]

An analogy with surrealism in art will help us see how this aspect of Wittgenstein’s philosophical method works. One of the functions of art is to provide a new interpretation of our surrounding world, or to uncover some hidden aspect of it. Surrealism accomplishes this by shifting objects from their familiar/unique setting and placing them in an unfamiliar/alien context; typically one that forms a contrast with their original home. The contrast might be one involving a shift from darkness to light, old to new, noise to silence, etc. For example, a surrealist work might involve moving a monastery from a craggy hill to the edge of the sea, or into a busy city, or some element from the busy city (say, the street with its pedestrians) onto a deserted island, etc. In this way the artist draws our attention to a given object -- an object we might have overlooked had it been in its familiar setting. He thereby sharpens our focus on it. This technique helps bring out ‘hidden’ aspects of the world, ‘hidden’ because they are so familiar that we overlook them; we take them for granted. They become insignificant from their familiarity. By rearranging phenomena, the artist makes the insignificant significant. Wittgenstein’s method of constructing hypothetical language-games can be viewed as an employment of a similar technique, only in a different medium: language. By taking us on journeys into ‘imaginary landscapes’ (imaginary uses of language), where words/sentences have a function that differs from their actual function, Wittgenstein helps us to see the familiar role we have overlooked: how they actually function. For it is forgetting this that gives rise to philosophical puzzlement. Wittgenstein’s method, as with surrealism, does not involve
building a new construction out of new material, but only ‘rearranging what we have always known’ [PI §109], like the ‘rearrangement of books in a library’. [BB p. 44]¹ Now, it is this re-arrangement of the familiar -- using familiar words in unfamiliar ways – that gives rise to a feeling of eeriness. Here are two examples to illustrate:

Now think of the following use of language: I send someone shopping. I give him a slip marked ‘five red apples’. He takes the slip to the shopkeeper, who opens the drawer marked ‘apples’; then he looks up the word ‘red’ in a table and finds a color sample opposite it; then he says the series of cardinal numbers – I assume that he knows them by heart – up to the word ‘five’ and for each number he takes an apple of the same color as the sample out of the drawer. – It is in this and similar ways that one operates with words. … [PI §1]

Let us imagine a language for which the description given by Augustine is right. The language is meant to serve for communication between a builder A and an assistant B. A is building with building-stones: there are blocks, pillars, slabs and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose they use a language consisting of the words ‘block’, ‘pillar’, ‘slab’, ‘beam’. A calls them out; -- B brings the stone which he has learnt to bring at such and such a call. – Conceive this as a complete primitive language. [PI §2]

Of course, the shopping expedition would not, in real life, be conducted in the manner portrayed in PI §1. No grocer keeps apples in drawers labeled ‘apples’ or consults color charts. We do not live in such ludicrous, mechanical worlds. The point is, we are supposed to contrast that imaginary language-game constructed in §§1,2 with our actual linguistic practices. In this way we might see (the contrast might help bring to light) that communication does not demand what not demand that every word must have something for which it stands -- that something being its meaning -- as the Augustinian picture (a pseudo-theory) assumes.

Conclusion

I have offered what I take to be four reasons for the cagey or eerie quality of Wittgenstein’s writing. They are, as we have seen, essentially tied with the aim/goal and operations of his philosophical method, and not, as many have held, merely decorative or inconsequential features of his philosophy.

¹Or the rearrangement of furniture in a room. According to Moore (1955) p. 27, Wittgenstein ‘compared his method to the tidying of a room where you have to move the same object several times before you can get the room really tidy’.
Bibliography


CHAPTER FIFTEEN

Collingwood and Rorty on the Role of Philosophy

Carol Nicholson

There are striking similarities between R.G. Collingwood and Richard Rorty in both their lives and their philosophical positions. As precocious children brought up in unconventional families, they received unusually broad educations and developed wide-ranging intellectual as well as other interests and talents. Early in life they were attracted to the Socratic concept of the philosopher as the one “who has a taste for every sort of knowledge and throws himself into acquiring it with an insatiable curiosity.” Both Collingwood and Rorty were influenced by Hegel and Marx and many other thinkers outside of Anglophone philosophy. They were both what we today would call “interdisciplinary,” emphasizing the central role of the arts and history as well as the sciences in informing the philosophical perspective. Although politically liberal, they were regarded as mavericks by the prevailing analytical philosophy of their times, and their rejection of the correspondence theory of truth provoked extremely hostile reactions among their contemporaries. They agreed that false dualisms in philosophy, especially the dichotomy between theory and practice, must be overcome, and their denial that there are any eternal philosophical problems or solutions resulted in the accusation that they had fallen into a dangerous relativism that would destroy the autonomy of philosophy.

A study of the relationship between R.G. Collingwood’s philosophy and Richard Rorty’s reinterpretation of American pragmatism is long overdue. In a review of Rorty’s Philosophy and the Mirror of Nature in 1981, Quentin Skinner pointed out that Collingwood anticipated several of Rorty’s arguments by mocking the idea of the philosopher as an “international boundary commissioner” in Speculum Mentis and challenging the Cartesian attempt to discover the indubitable foundations of knowledge in An Essay on Metaphysics. Skinner found it surprising that Rorty never mentions Collingwood’s work, since Collingwood made “precisely the arguments” mounted by Rorty’s heroes, Dewey, Wittgenstein, and Heidegger. Parallels

4 Ibid., 46.
between Collingwood’s philosophy and pragmatism were pointed out as early as 1969 by Louis Mink,¹ who noted that the first sentence of Collingwood’s *Speculum Mentis* is “All thought exists for the sake of action” and the final chapter of his *Autobiography* is entitled “Theory and Practice.” Gary Ciocco recently wrote, “I personally hope someone can meet Mink’s decades-old challenge—not yet attempted let alone met, as far as I know—to prove Collingwood part of the same philosophical stream”² as the pragmatists. This paper will take a first step towards meeting the challenge of exploring the relationship between these two schools of thought. Although I do not go so far as to conclude that Collingwood is “part of the same stream” as pragmatism, I argue that by a different route he ended up in a similar position on the role of philosophy.

In the “Preface” to the most recent volume of his collected works, *Philosophy as Cultural Politics*,³ Rorty wrote that these papers, like his previous writings, are “attempts to weave together Hegel’s thesis that philosophy is its time held in thought with a non-representationalist account of language,” arguing that “Hegelian historicism and a Wittgensteinian ‘social practice’ approach to language complement and reinforce each other.”⁴ These two theses were also central to Collingwood’s work and were the reason for his intellectual isolation at Oxford and his kinship with the Italian philosophers, Croce, Gentile, and Ruggiero, whose long familiarity with the tradition of Vico’s historicism and non-representational view of language enabled them to be more sympathetic to Collingwood’s ideas than were most British philosophers at the time. A comprehensive comparison demonstrating that Rorty and Collingwood were in agreement in accepting Hegel’s historicism and a “social practice” view of language is beyond the scope of this paper. My analysis is intended to provide a preliminary guide to translating Rorty’s pragmatic idiom into Collingwood’s philosophical vocabulary, focusing on what they thought philosophy should not be, what they thought it should be, and how they thought philosophy should be related to other disciplines of knowledge.

The attempt to follow up on the suggestions of Skinner, Mink, and others to find parallels between Collingwood’s philosophy and pragmatism encounters many obstacles. For one thing, there is a widespread tendency to think of pragmatism as a distinctly “American” kind of philosophy. One interpretation along these lines is that pragmatism is an expression of the frontier mentality of the pioneer town,⁵ and another views it as a reaction

---

⁴Ibid, ix.
against the ideological certainties of the Civil War.¹ Both of these approaches may be useful for understanding the mentality of the “Metaphysical Club” from which pragmatism in America arose, but the notion of pragmatism in general as an expression of the American spirit fails to take account of British pragmatism, represented by F.C.S. Schiller, and the schools of pragmatist thought that flourished in Italy in the early twentieth century. Replacing the tendency to claim American ownership of pragmatism with a more cosmopolitan approach that includes both American and non-American philosophers will allow for an unbiased and open-minded comparison between Collingwood and pragmatism.²

Another difficulty in attempting to find common ground between Collingwood and the pragmatists is the need to look beyond Collingwood’s own words. There are very few references to pragmatism in Collingwood’s writings, and they express an extremely negative attitude toward it. He calls James’ Varieties of Religious Experience “a fraud,”³ and in Speculum Mentis he refers to the “babblings of pragmatism.”⁴ There is no evidence of any influence on Collingwood from British, American, or Italian pragmatists either in his published writings or his unpublished manuscripts, and any of his views that resemble pragmatism were arrived at independently. Differences in terminology make Collingwood’s position appear more dissimilar to pragmatism than it actually is. For example, Collingwood never used the terms “anti-foundationalism” and “fallibilism,” but his thought clearly exemplifies these positions as the terms are now understood. What he calls the “logic of question and answer” closely resembles the pragmatic theory of inquiry and appears to espouse what we would now call a “pragmatic theory of truth,” although he would quite probably have rejected this characterization of his thought, because his understanding of the new philosophy of pragmatism was filtered through Croce’s negative view of the Italian pragmatists.⁵

Collingwood’s use of the term “metaphysics” to describe the study of the presuppositions of science was misleading in its supernaturalist connotations, since his intention was to develop a conception of philosophy that was as thoroughly naturalistic as that of Dewey. The expression “absolute presuppositions” was also extremely unfortunate in that there is nothing

³Collingwood, An Autobiography, 93.
“absolute” about them except for people’s stubborn resistance to having them challenged, comparable to the “method of tenacity” described by C.S. Peirce in “The Fixation of Belief,” which Collingwood does not appear to have known. Further compounding the problem of comparison, Rorty’s language is sometimes as misleading as Collingwood’s, as in his notorious expression, “edifying conversation,” as a description of a new self-image for philosophy, connoting that moral or spiritual improvement, as opposed to truth, is the goal of philosophy. My view is that both Collingwood and Rorty were attempting to create new vocabularies where no words yet existed to express their remarkably similar concepts of the role of philosophy. Although some of their choices of terminology have not been successful in persuading their immediate audiences, looking beyond their misleading forms of expression reveals interesting parallels and may shed light on both of these bold and imaginative visions.

According to Stephen Toulmin, Collingwood was “one of those English nonconformists who are fated to end up by creating a party of one member.”1 In his autobiography Collingwood described his position with respect to the dominant philosophical trends in Oxford during the 1920’s and 1930’s by saying that he was “cut off not only from the ‘realist’ school to which most of my colleagues belonged, but from every other school of thought in England, I might almost say in the world.”2 His intellectual isolation resulted, at least in part, from the fact that he was a systematic philosopher, not in the sense that his thought formed a comprehensive system of thought like that of Hegel or Spinoza, but in the sense that he wrote on all of the main topics in philosophy and had a methodical plan for broadening its scope by outlining its relationship to other disciplines, including art, religion, science, and history. During a period in which philosophers were reacting against Hegelian idealism and favored a piecemeal approach to the solution of philosophical problems, Collingwood’s synthetic habit of mind often caused his work to be misinterpreted or dismissed, in the words of A.J. Ayer, as “the debris of Absolute Idealism.”3 He vehemently denied the label, however, and wrote in correspondence with Gilbert Ryle, “Why presume me an Idealist? I have nowhere . . . so described myself. . . I represent nobody but myself and ‘protest’ no particular allegiance to ‘Plato, Kant in his less Humean moods, and Hegel; in fact I protest allegiance to nobody.”4

Louis Mink wrote that “it is useless to ask whether Collingwood was, at least in intention, a ‘realist’ or an ‘idealist’, because he regarded both realism and idealism as partial views which need to be corrected by being absorbed into a more comprehensive theory.”5 Mink aptly described Collingwood’s

1Stephen Toulmin, “Introduction” to Collingwood, An Autobiography, x.
5Mink, Mind, History, and Dialectic, 112.
philosophy as “a continuous examination of the possibility and nature of
dialectical thinking,” which has affinities with both pragmatism and
existentialism, while admitting that Collingwood himself would no doubt have
eschewed those labels as well.1 Despite his lack of followers during his
lifetime, Collingwood’s reputation has risen during the past 60 years, and
Simon Blackburn’s entry in the Routledge Encyclopedia of Philosophy
describes Collingwood as “the greatest . . . philosopher of history . . . and one of
the greatest polymaths of twentieth century British philosophy.”

Although Collingwood’s contributions to the philosophy of history are now
widely recognized, it is less well known that he also wrote extensively on art,
religion, science, ethics, and politics throughout his career, beginning with his
first published book, Religion and Philosophy and ending with The New
Leviathan or Man, Society, Civilization, and Barbarism and the posthumously
entitled “The Place of Art in Education”2 Collingwood wrote, “Neither
education, nor politics, nor religion, nor anything else, can be improved. . .
[except] by a more penetrating vision of its proper and fundamental nature and
a more sustained effort to realize this nature in everyday life.”3 The proper
function of philosophy, he maintained, is not to engage in abstract thought for
the sake of advancing “pure” knowledge, but to make people’s lives better by
providing them with a clearer understanding of what they are trying to
accomplish. This essay laid down the pattern for Collingwood’s attempt to
integrate all of the branches of knowledge in line with the theme of restoring
the “unity of mind,” which he had already emphasized in “Ruskin’s
Philosophy” in 1919 and Speculum Mentis in 1924. He had discovered, at
about the age of nine, by running across a 17th c. book (probably Descartes’
Principia) at a friend’s house, the “secret” modern textbooks tend to keep
hidden that “science is less like a hoard of truths, ascertained piecemeal, than
an organism which in the course of its history undergoes more or less
continuous alteration in every part.” He later worked out the general
implications of this insight into the principle of all forms of intellectual
endeavor: No work of art or literature or science is ever “finished”; the effort of
improving it ceases only because deadlines are at hand, energy and inspiration
fail, or problems remain intractable.4 He argued that history is not the
“scissors-and-paste” chronicling of dead facts, but it is the activity of
rethinking in the present the thoughts that inspired the actions of people in the
past by imagining the circumstances in which they worked out their problems.
He thought that the proper role of philosophy should not be providing (or
deconstructing) justifications for propositions or finding solutions to “eternal
problems” such as how to reconcile mind and body or freedom and

1Mink, Mind, History, and Dialectic, 7–12; 16; 112.
reprinted in Alan Donagan, ed. Essays in the Philosophy of Art (Bloomington: Indiana
University Press, 1964),
3Donagan, ed. Essays in the Philosophy of Art, 188.
4Collingwood, An Autobiography, 2.
determinism, but rather the attempt to give people a clearer understanding of what they are doing when they go about their everyday business of living and thinking and solving the problems that arise in human experience.

In *The New Leviathan* Collingwood outlines his theory of civilization as a complex combination of processes by which the members of a community 1) become more able to control the natural world intelligently in order to maintain and ameliorate their lives; and 2) learn to treat other human beings with civility rather than with force, respecting both their own freedom of will and that of others with whom they associate, both within and outside their community. The mainspring of the process by which people become civilized and maintain their civilization is a cooperative attitude towards knowledge. Collingwood contrasts what he calls an “eristic of knowledge, a tendency to make it a matter of contention and competition and monopoly,” with a “dialectic of knowledge, a tendency to make it a matter of agreement and cooperation and sharing.”

Obviously, no civilization can last more than a single generation without a “dialectical” willingness to teach and to learn, and philosophy is not required in order for the normal process of civilization to take place. But philosophy does have a role to play, Collingwood believed, if there is confusion and doubt about what civilization means.

In his autobiography Collingwood wrote, “My life’s work hitherto, as seen from my fiftieth year, has been in the main an attempt to bring about a *rapprochement* between philosophy and history . . . I was also working at a *rapprochement* between theory and practice.” In a letter to the editor of *Philosophy* in 1934, Collingwood wrote, “As the seventeenth century needed a reasoned conviction that nature is intelligible and the problems of science in principle soluble, so the twentieth needs a reasoned conviction that human progress is possible and that the problems of moral and political life are in principle soluble. In both cases the need is one which only philosophy can supply.” Collingwood believed that the role of the philosopher is not like that of the pilot of a ship who can descend like a *deus ex machina* and dictate to the members of the crew how to navigate around the rocks and shoals of life; nor, on the other hand, should the philosopher be a mere spectator, disinterestedly theorizing about the concepts of navigation from an ivory tower above the shore. The truth, he thought, lies between these two extremes. The philosopher is one of the crew whose special and essential function is not to solve the detailed problems of education, morality, economics, and politics, but rather to give “a reasoned statement of the principle that there can be no evils in any human institution which human will cannot cure.”

Collingwood pointed out that Alexander, Whitehead, and other philosophers of science in his time initiated a new movement in which the epistemological controversies between realists and idealists fell into the background and the central topics of discussion were the ideas of evolution, development, history,

---

Collingwood and Rorty on the Role of Philosophy

and progress. From this new perspective, which attempted to integrate the insights of a variety of disciplines, “philosophy feels itself a collaborator with science, neither its enemy nor its slave, but having its own dignity and its own methods, while it respects those of science.” Collingwood’s vision of philosophy as a collaborator with history, science, and the other disciplines was also intended to provide a rapprochement between theory and practice by getting philosophers back in the business of talking to ordinary people about the vital practical issues of their generation. His autobiography recounts the gradual realization that the kind of philosophy that was being taught at Oxford was having devastating effects upon students. “The pupils, whether or not they expected a philosophy that should give them . . . ideals to live for and principles to live by, did not get it; and were told that no philosopher . . . would even try to give it. . . If the realists had wanted to train up a generation of Englishmen and Englishwomen expressly as the potential dupes of every adventurer in morals or politics, commerce or religion . . . no better way of doing it could have been discovered.” In 1939 Collingwood explicitly connected the failure of philosophy to perform its proper role with the rise of Fascism by writing, “The fate of European science and European civilization is at stake. The gravity of the peril lies especially in the fact that so few recognize any peril to exist. When Rome was in danger, it was the cackling of the sacred geese that saved the Capitol. I am only a professorial goose, consecrated with a cap and gown and fed at a college table; but cackling is my job, and cackle I will.” His autobiography concludes, “I know that all my life I have been engaged unawares in a political struggle, fighting against [the end of clear thinking and the triumph of irrationalism] in the dark. Henceforth, I shall fight in the daylight.”

In spite of the differences between the starting points and methods of Collingwood and Rorty, there are clear parallels between the visions of philosophy at which they arrived. Rorty was not a “systematic” philosopher in any sense of the word, and his reputation was that of a “system-destroyer” rather than a “system-builder,” but he was known for his wide reading in both historical and analytic philosophy, his synoptic bent that was unusual among his colleagues, and his concern that professional philosophers had become isolated from the rest of culture. According to Eduardo Mendieta, Rorty was “probably the most interdisciplinary philosopher that the United States has produced.” Writing on law, religion, science, and literature, he was “the most ecumenical” American philosopher of the last half century and was “incredibly cosmopolitan in his approach to philosophical issues.”

Rorty’s strategy in his first major work, Philosophy and the Mirror of Nature, was to deconstruct the whole epistemological tradition by undermining its central metaphor, arguing that the visual imagery used to describe

---

1Collingwood, An Autobiography, 48—49.
knowledge has led philosophers astray ever since Plato’s allegory of the cave. Rorty recommended abandoning the idea that philosophy is a sort of “master discipline” with privileged access to the mirror image of correspondence with reality, and replacing this image with the idea of philosophy as “edification,” or “hermeneutics,” which would be constrained only by the concern to keep the conversation going, rather than submitting to something transcendent, big, and powerful, such as God or Truth. In *Consequences of Pragmatism* Rorty argued that if we abandon epistemology and “uppercase” Philosophy—traditions that he thought had outlived their usefulness—we can engage in “lowercase” philosophy, defined more modestly in Wilfred Sellars’ phrase, as the attempt to “see how things, in the broadest possible sense of the term, hang together, in the broadest possible sense of the term.”¹

In later works, such as *Contingency, Irony and Solidarity, Philosophy and Social Hope*, and *Philosophy as Cultural Politics*, Rorty appealed to the positive goals of intersubjective agreement and solidarity as substitutes for grounding theories in something transcendent such as Reality or God. His provocative views shocked philosophers on the right, who were afraid that the denial of the correspondence theory of truth would destroy the basic presuppositions of Western civilization, as well as those on the left, who thought that his denial of the foundational role of philosophy would undermine the project of overthrowing capitalism and male supremacy. On the other hand, one former student wrote, “For many of us [Princeton University philosophy majors], Rorty functioned as a truth teller, an ironic role for a thinker who became known as an ‘ironist’ skeptical of truth.”² According to Robert Brandom, in spite of the common criticism that Rorty was a relativist who did not take philosophical problems seriously, “Nothing could be further from the truth. . . He took them as seriously as it’s possible to do. It’s just that, taking them seriously, he was still not willing to kid himself about them.”³ Rorty eventually adopted the slogan that “if you take care of freedom, truth takes care of itself. A true statement is just one that a free community can agree to be true. If we take care of political freedom, we get truth as a bonus.”⁴

In spite of the fact that many philosophers criticized Rorty for appearing to try to put an end to the discipline, not since John Dewey has America had a more enthusiastic cheerleader for the wider role of philosophy in public affairs. Towards the end of his life, Rorty’s project was to expand philosophy’s range and make it more relevant to people’s everyday lives by reaching an audience beyond professional philosophers. He carried out this project with an astonishingly large number of interviews, speaking engagements, and writings, including articles in the *New York Times*, the *Chronicle of Higher Education, The Nation, Dissent*, and countless other popular magazines, in addition to

---

continuing to publish scholarly books and articles. Rorty’s introduction of “she” as the generic pronoun, much debated a generation ago, has now been widely adopted, but few have acknowledged his instrumental role as President of the American Philosophical Association during the pluralist revolution in 1979, which has had a lasting effect on the American philosophical profession.1

Rorty committed himself relatively early in his career to the pragmatic intellectual position that theory and practice are inseparable, but like Collingwood, he took a while to come to the practical realization that he could not keep his philosophy and his politics in separate corners. In 1985 Rorty announced at the Eleventh Inter-American Congress of Philosophy in Guadalajara Mexico that “we should not assume it is our task, as professors of philosophy, to be the avant-garde of political movements.” By 1998, with the publication of his first overtly political work, Achieving Our Country,2 he had come to see that “the natural consequences of his own worldview required engagement in ‘philosophy as cultural politics,’ the title he chose for the fourth and final volume of his papers from Cambridge University Press.”3 Rorty eventually adopted Dewey’s view that “the history of philosophy is best seen as a series of efforts to modify people’s sense of who they are, what matters to them, what is most important” as well as Dewey’s hope that philosophy professors would see “intervention in cultural politics as their principal assignment.”4 Although he regarded the professionalization of philosophy and its transformation into an academic discipline a necessary evil, he thought philosophy should resist attempts to make it into an autonomous quasi-science. “The more philosophy interacts with other human activities—not just natural science, but art, literature, religion, and politics as well—the more relevant to cultural politics it becomes, and thus the more useful. The more it strives for autonomy, the less attention it deserves.”5 According to Carlin Romano, “‘To teach how to live without certainty,’ Russell wrote, ‘and yet without being paralyzed by hesitation, is perhaps the chief thing that philosophy, in our age, can still do for those who study it.’ Richard Rorty did that—magnificently and magisterially.”6

Like Collingwood, Rorty had premonitions of the end of democracy and tried to sound a warning. In an article in The Nation in October 2002, five months before the invasion of Iraq, he correctly predicted that Democratic Congressmen would grant President Bush the authority to wage the war out of

---

5Ibid., x.
fear of being depicted as “effete,” observing that “we may be able to keep the moral gains—the increases in political freedom and in social justice—made by the West in the past two centuries even if 9/11 is repeated year after year. But we shall only do so if the voters of the democracies stop their governments from putting their countries on a permanent war footing—from creating a situation in which neither the judges nor the newspapers can restrain organizations like the FBI from doing whatever they please, and in which the military absorbs most of the nation’s resources.” More recently, he said in an interview, “The Bush Administration has now been repudiated by U.S. public opinion, and the Iraq debacle will make future European governments hesitant about following America’s lead. But I still think that the end of democracy is a likely consequence of nuclear terrorism, and I don’t know how to guard against this danger. Sooner or later some terrorist group will repeat 9/11 on a much grander scale. I doubt that democratic institutions will be resilient enough to stand the strain.”

By different paths, Collingwood and Rorty ended up in similar places with regard to their views of the proper role of philosophy. Both challenged the separation of philosophy from other disciplines of knowledge and from practical action. They shared the view that philosophy is not the master-discipline, the Queen of the Sciences, or the international boundary commissioner which should resolve disputes among the disciplines about which propositions correspond most closely to the truth. They believed that philosophy’s job is to articulate clearly the meaning of the practices that people engage in when they pursue artistic, religious, historical, ethical, and political activities, helping to provide them with a more self-conscious understanding of their identities as individuals and as members of social and political communities. They envisioned philosophy as engaging in a collaborative activity with the arts and sciences and were committed to the idea that an inseparable part of philosophical understanding involves acting upon it, especially resisting impositions upon the free exercise of the ongoing pursuit of truth.

A different conclusion is reached in Giuseppina D’Oro’s *Collingwood and the Metaphysics of Experience,* which presents Collingwood and pragmatism as opposed on several important issues. D’Oro claims that while Collingwood took the perennial problems of philosophy seriously, the pragmatists tried to dissolve them by showing that they are meaningless. This interpretation seems to me to be wrong-headed on two counts. First of all, it cannot be squared with Collingwood’s statement in *An Autobiography* that the idea of “eternal problems” in philosophy is “a vulgar error, consequent on a kind of historical myopia which, deceived by superficial resemblances, failed to detect profound differences.” Secondly, it fails to acknowledge the constructive resolutions to traditional philosophical problems that James and Dewey arrived at by

---

overcoming what they perceived to be “false dualisms” such as mind and body. D’Oro maintains that pragmatists, unlike Collingwood, deny the role of norms in philosophy, but she seems to misunderstand the pragmatic reinterpretation of norms, which Rorty describes by saying that norms are not a priori truths but summaries, abbreviations, or rhetorical ornaments of human practices, rather than their foundations. On my reading, Collingwood is in agreement with the pragmatists that theory is not prior to practice, but philosophical principles are discovered by reflection on the norms that are implicit in practical experience. In an otherwise excellent book, it is unfortunate that D’Oro misinterprets both Collingwood and pragmatism on these points.

If my argument has been successful in showing that there are strong similarities between Collingwood and Rorty, two conclusions seem to follow. First of all, far from being at the dead end of British idealism as Collingwood is often interpreted, he was ahead of his time in anticipating the redescriptions of philosophy and its role in relation to other disciplines that Rorty presented. Secondly, Rorty’s iconoclastic vision of the role of philosophy has roots not only in American pragmatism but also in the broader philosophical tradition upon which Collingwood’s writings were based. What Rorty calls the “edifying conversation” of humanity and Collingwood calls the “dialectical” spirit of agreement that underlies the social contract can be viewed as different ways of trying to describe the ongoing and endless quest for wisdom that Socrates had in mind when he defined the philosopher as the one “who has a taste for every sort of knowledge and throws himself into acquiring it with an insatiable curiosity.” Rorty wrote, “James and Dewey were not only waiting at the end of the dialectical road which analytic philosophy traveled, but are waiting at the end of the road which, for example, Foucault and Deleuze are currently traveling.” I have argued that Collingwood has been waiting for nearly a century at the end of the road that Rorty was traveling, but it was Rorty who succeeded in bringing some of Collingwood’s ideas to the vanguard of philosophical debate. Both Collingwood and Rorty made themselves unpopular by asserting that the division of labor in the academy had gone too far in making a distinction into a sharp separation between philosophy and the other disciplines. My own view is that (on this point, at least) they were right, and we should carry on an important part of their legacy by attempting to integrate philosophy with all branches of knowledge in the spirit of Socrates and by continuing the struggle to make the world safe for the broad definition of philosophy that they envisioned.

2Rorty, Consequences of Pragmatism, xviii.
3My view of the role of philosophy is greatly indebted to Guy W. Stroh, my former Department Chair at Rider University, who often said that our mission is to “make the world safe for philosophy.”
Even thought Richard Rorty was not a philosopher of education, he had inclinations to education being an important field of action and change. Among his works, some have more educational relevance. First, his *Philosophy and Mirror of Nature* (1979) should be mentioned in the final chapter of which he suggests the concept of 'edification' as an alternative for education. Edification is interesting for him because it implies that people can free themselves from the dominance of current normative discourses and communicate with other paradigms to find out new and better ways of talking.

Second, his view on contingency of language (Rorty 1989) is worth mentioning because of its important consequences for educational matters. Referring to Donald Davidson's view on metaphor, he regards it as a procedure for creating new ways of seeing the world and human relations. According to Rorty, this kind of basic change and challenge could be expected in tertiary education in university being a suitable ground for individuality.

Distinction between two main phases of education, before and after entering university, is what Rorty pays attention to in his article titled *Education without Dogmatism* (1989). Here he deals more explicitly with education and demarcates two phases for it; one being the realm of socialization and the other, namely university, a period for individuality and ironic treatment with the established thoughts and views. Rorty's (1994) interpretation on feminism in terms of pragmatism also shows a general view on action and education.

In what follows, I will give a brief account of Rorty's important educational views along with an examination and challenge with some of their implications.

### The Nature of Education

In the first part, I will deal with Rorty's view on the nature of education in terms of edification. Then, I will have a critical discussion on his view in relation to indoctrination.
Education as Edification

Rorty is in full agreement with Gadamer (1989) in his challenge with method, as a clear way to truth, as well as with his revolutionary treatment with the longstanding philosophical tradition in the West.

As for the first point, Rorty regards Gadamer comparable to Feyerabend (1978). Accordingly, Feyerabend tried to reject the tight rationality of positivism by his famous 'Anything Goes', as Gadamer intended to reject the hermeneutical attempts (like that of Dilthey) in order to follow the methodological rigor of natural sciences in humanities (Rorty, 1979, p. 358).

Concerning the second point, Rorty takes Gadamer to play a significant role in providing a paradigmatic distance from the long tradition of philosophy in the West. This is because, according to Rory, while after Plato the main aim of philosophy has been 'knowledge', Gadamer introduced Bildung as the aim of thinking. Rorty is not, of course, certain about Plato's position with regard to the idea of human as knower. Thus, he states that after millenniums of commentary on Plato's dialogues, it is not still clear whether he was a systematic revolutionary philosopher or an edifying revolutionary one (ibid, p. 369). Rorty introduces 'education' and 'self-formation' as candidates for translation of Bildung, but finally prefers 'edification' for it. Thus, according to him, Gadamer belongs to the camp of few new philosophers who have tried to shift the philosophical paradigm of human as knower:

In the main stream of the Western philosophical tradition, this paradigm has been knowing—possessing justified true beliefs…on the periphery of the history of modern philosophy, one finds figures…distrust of the notion that man's essence is to be a knower of essences. Goethe, Kierkegaard, Santayana, William James, Dewey, the late Wittgenstein, the later Heidegger, are figures of this sort. (ibid, pp. 366-7)

If we ask, what these philosophers were seeking to replace for 'knowing', Rorty will answer by 'edifying' which indicates, in turn, using the practical wisdom. Thus, when he talks of 'wisdom as something of which the love is not the same as that of argument', he considers it 'as the practical wisdom necessary to participate in a conversation'. (ibid, p. 372)

In clarifying edification in terms of practical wisdom, Rorty does not mean what, for instance, Kant meant by the term demarcated from the pure reason. This is because Rorty thinks that such distinctions, like the positivist distinction between fact/value, presuppose the representational mind and seeking the truth. However, what Rorty looks for is to put edification at the center and regard the factual discourse as just one kind of edification. In this way, he is in full agreement with Gadamer: 'This is why Gadamer devotes so much time to breaking down the distinctions which Kant made among cognition, morality, and aesthetic judgment.' (ibid, p. 364)
To consider the factual discourse of empirical sciences as just one feature of edification requires that education is not reduced to instruction. Again he accompanies Gadamer, against Mill and Carnap, in fending off the demand for objectivity in the *Geisteswissenschaften* and considers this attempt as to 'prevent education form being reduced to instruction in the results of normal inquiry.' (ibid, p. 363)

In fact, Rorty not only rejects the attempts to reduce education to instruction, but also wants to reduce instruction to education. In other words, he wants to say that instruction of facts, being provided by the normal inquiry, should be considered as just one feature of a broader human attempt namely edification. While fact/value distinction renders value to a non-cognition matter, Rorty's attempt is to render the factual discourse to a feature of value domain. That is why he says that the final result of the attempts of admirers of empirical method and science, like Peirce, Dewey, and Popper, is solely some ethical values, namely those of the open society (Rorty, 1999, p. 36).

Edification being so central in human activity and thought has, in addition, a linguistic feature. It is, in fact, a kind of description of things and human situations that should be assessed in terms of its effectiveness for coping rather than in terms of representation of reality. Referring to this linguistic aspect of edification, Rorty says: 'I shall use "edification" to stand for this project of finding new, better, more interesting, more fruitful ways of speaking.' (1979, p. 360) The words 'better' and 'more fruitful' indicate his pragmatic concern in coping rather than representing reality.

A further characteristic of edification relies on its intercultural and interdisciplinary dimension. In seeking 'better' ways of speaking, it might be necessary to edify one's own cultural or disciplinary limitations by means of incorporating other cultures' or disciplines' strengths. Rorty considers this as a hermeneutic activity for making connections between our own culture and some exotic culture, or between our own discipline and another discipline. While these cultures or disciplines might seem in the first place to pursue incommensurable aims in terms of incommensurable vocabularies, the 'hermeneutic activity' can in principle be expected to remove the barriers. (ibid)

Rorty reserves a further way for edification by means of 'poetic activity' being in inverse direction to the hermeneutic activity. Using Davidson's (1996) conception of metaphor, he regards this way as a metaphorical or poetic way of thinking. Davidson criticized the current conception of metaphor for reducing metaphorical meaning to the literal meaning with the only difference of being hidden. Instead, he holds that metaphor is indeterminate in meaning and, thus, can endlessly be created again and again. Referring to this kind of edification and its inverse direction compared to hermeneutic activity, Rorty believes that it should be 'followed by, so to speak, the inverse of hermeneutics: the attempt to reinterpret our familiar surroundings in the unfamiliar terms of our new inventions.' (1979, p. 360) According to him, this way of edification, being ironical and radical in nature, can only be expected in higher education.

Education as edification needs to be preceded by acculturation or socialization. There must be something that can be regarded as the subject of
edification and that is why edification can not be the starting point. According to Rorty, to start education with acculturation is perhaps a necessary background for edification and this will be the case even for the education of the revolutionary or the prophet.

I raise this banal point that education—even the education of the revolutionary or the prophet—need to begin with acculturation and conformity merely to provide a cautionary complement to the 'existentialist' claim that normal participation in normal discourse is merely one project, one way of being in the world. (ibid, p. 365)

Having explained the main characteristics of edification, I am going now to draw and criticize its implications.

**Blurred Borders of Education and Indoctrination**

Rorty's reliance on acculturation as part of or preparation for education led him to resist any conception of education incorporating argument or discussion as a constitutive element like the definitions inspired by Habermas's (1984) view on communicative action. In order to distinguish between argument and education, Rorty says:

Let me use the distinction between *arguing* with people and *educating* people...If all education were a matter of argument, this distinction would collapse. But, unless one broadens the term 'argument' beyond recognition, a lot of education is not. In particular, a lot of it is simple appeal to sentiment. (2002, p. 19)

This way of equating sentiment, void of discussion and argument, with education reminds us of the well-known dichotomy of emotion and reason in the realm of education. Rorty is, of course, right in claiming that education should not be limited to argument and that sentiment is also an important dimension of education. However, he is not quite right in drawing a clear-cut distinction between sentiment and argument. The combination of sentiment and discussion is what to be sought in education even though with different degrees on a continuum. To put in Kantian terms, sentiment without argument is blind and argument without sentiment is empty.

Rorty's view on sentiment blurs, at the same time, the longstanding distinction philosophers of education (e.g. Hirst, 1974) made between indoctrination and education. It is undoubtedly true that edification is logically and/or empirically preceded by a period of conformity or acculturation in the early childhood. Given Rorty's attempt to equate education with edification, a better strategy would be to consider childhood as the preparatory period for education (edification) rather than being constitutive of it. In other words, one could say that education as edification does not start until some culture is
absorbed by the child. In this sense, acculturation is not education *per se* but an inevitable background for it to occur.

However, it seems that Rorty does not want to confine conformity and acculturation to the period of childhood but to consider them absolutely precedent to edification even in the case of adults who have already absorbed some culture. This point shows that Rorty not only blurs the border of education and indoctrination, but also embraces this blur to provide a particular conception of education. This is evident in his recourse to the statement of the victorious Allied armies in Germany and Japan which is worth quoting here despite its length:

> It seems to me that the regulative idea that we – we wet liberals, we heirs of the Enlightenment, we Socratists – most frequently use to criticize the conduct of various conversational partners is that of "needing education in order to outgrow their primitive fears, hatreds, and superstitions." This is the concept the victorious Allied armies used when they set about re-educating the citizens of occupied Germany and Japan. It is also the one which we used by American schoolteachers who had read Dewey...It is a concept which I, like most Americans who teach humanities or social science in colleges and universities, invoke when we try to arrange things so that students who enter as bigoted, homophobic, religious fundamentalists will leave college with views like our own. (Rorty, 2002, p. 21)

Here, Rorty uses education with the function of making people similar in their viewpoints. That is why he uses the phrase 'we try to arrange things so that' indicating that this kind of education is merely emotional without the ingredients of argument and discussion. It is very strange to hear from Rorty that he negates even 'the possibility of reformulating' his views before the fundamentalists:

> When we American college teachers encounter religious fundamentals, we do not consider the possibility of reformulating our own practices of justification so as to give more weight to the authority of the Christian scriptures. Instead, we do our best to convince these students of the benefits of secularization. (ibid, p. 22, Emphasis added)

> It seems that by rejecting the possibility of reformulating his views, Rorty, in fact, rejects some logical consequences of his own views. In his above-mentioned clarification of edification, he stated that "the attempt to edify (ourselves or others) may consist in the hermeneutic activity of making connections between our own culture and some exotic culture" (1979, p. 363) This statement requires him to accept the possibility of reformulating his views to see and verify their possible strengths and weaknesses; otherwise, it would
lead to a kind of dogmatism, while he looks for 'Education without Dogmatism' (Rorty, 1989b). What makes his statement still stranger is that he talks at the position of a college teacher, while in the same article he considered education in university as a period for individuality and irony toward any kind of final vocabulary.

Rorty sees his way of confrontation with such students as the best way, while considers the admirers of transcendental reason and discussion, like Putnam and Habermas, unable to handle them:

But I think that the handling of such students is a problem for Putnam and Habermas. It seems to me that I am just as provincial and contextualist as the Nazi teachers who made their students read *Der Stürmer*; the only difference is that I serve a better cause. I come from a better province. I recognize, of course, that domination-free communication is only a regulative ideal, never to be attained in practice. (Rorty, 2002, p. 22)

When Rorty claims that the only difference between him and the Nazi teachers is that he comes from a 'better' province, then one might ask whether this judgment of being 'better' refers merely to an inner psychological sentiment or there is public evidence to show why it is better. If it is merely an inner sentiment, then there are such sentiments within the Nazi teachers about their provinces as well. And if there is public evidence to consider, then Rorty has to follow the way that Putnam and Habermas suggested in terms of a reasonable discussion. If so, then Rorty should not distinguish between the education of fundamental parents of their children and discussion and, thus, should not try to make their views seem silly rather than discussable. Referring to those parents, he says:

You have to be educated in order to be a citizen of our society, a participant in our conversation, someone with whom we can envisage our horizons. So we are going to go right on trying to discredit you in the eyes of your children, trying to strip your fundamentalist religious community of dignity, trying to make your views seem silly rather than discussable. We are not so inclusivist as to tolerate intolerance such as yours. (ibid, p. 22)

Even though Rorty is so intolerant and exclusivist with regard to the non-members of his society, he looks for solidarity and consensus among its members. This brings us to a further dichotomy he draws between objectivity and solidarity.
Rorty's Unnecessary Dichotomies

In parallel to his rejection of representationalism, Rorty (1991, 1989a) rejects objectivity as an ideal in science and, hence, in science education. In the same vein, he undermines epistemological trend of looking for research methods and replaces it with conversation, as replaces objectivity with solidarity.

Referring to the central position of conversation in inquiry as a requirement of pragmatism, Rorty defines pragmatism as 'the doctrine that there are no constraints on inquiry save conversational ones' (1982, p. 165). By saying this, he means there are no constraints for inquiry, as might be supposed, due to the nature of the objects studied, or the minds, or the language itself. The only reasonable constraints, according to him, are the ones rooted in our conversation with our fellow-inquirers.

This insistence on the central position of conversation in science and inquiry shows Rorty's reliance on Thomas Kuhn's paradigmatic view, particularly in the phase of normal science in which consensus of scientists draws the boundaries of science and the scientific. Thus, Rorty makes a sharp distinction between conversation and research and thereby limits reasonable constraints to the former. So far as consensus is concerned, it is also significant not to neglect the importance of abnormal discourse in the development of scientific findings. As Ruth (1987) states, according to people like Feyerabend, if we understand the role that innovative reasoning plays in scientific findings, then 'this is sufficient to establish that abnormal discourse is a pervasive and integral part of the research process.' (p. 107)

At a deeper level, we need to think about the dichotomy between objectivity and solidarity or consensus. Rorty's attempt to avoid representational concerns is, of course, a reasonable endeavor. It is reasonable and vital in so far as preventing us from taking our conjectures as the picture of reality. However, he goes too far in cutting the relation between science and reality and considering science as a linguistic device with a solely pragmatic and instrumental significance. Others, like Popper (1959), took the concerns of conflating conjectures with reality serious and, at the same time, more reasonably reserved a place for truth as a final regulative rule for scientific activities.

In fact, in order to justify why our solutions work in relation to the world, we cannot avoid, in the final analysis, some kind of representational account. It is, of course, evident that solving a problem does not necessarily indicate a correspondence with reality, as it is clear in the case of successful anticipations made in the past based on the theory of flat earth. However, if one asks why such wrong theories work, the answer will contain some kind of representational account in the final analysis. Referring to successful wrong theories, Rescher (1987, p. 75) talks about the Nature's tolerance. That is to say, a wrong theory works because its anticipations are within the range of Nature's tolerance. For instance, architectures can rely, even at the time being, on the theory of flat earth to construct a building, but this theory works because the mistakes are negligible compared to the curve of the earth. However, it would
not be possible to construct a building on a surface as wide as a quarter of the earth.

Thus, the dichotomy of objectivity and solidarity does not seem reasonable because solidarity being sought in the realm of problem solving rests, in the final analysis, on objectivity. Although we should admit all the reasonable cautions in relating our scientific activity to reality in order to avoid naïve interpretations, nevertheless we should not dismiss objectivity from our scientific endeavor altogether. In consequence, it is not tenable to substitute consensus and conversation for objectivity in science education.

Conclusion

Relying on the insights of both analytic and continental traditions in philosophy, Rorty reconstructed pragmatism and its educational implications. While the heir of analytic philosophy presses on science and its methods, continental philosophy puts emphasis on culture and the lived experience. Rorty takes science important not for its methodological rigor but because of its ability to provide consensus in society. Consequently, in the realm of education, he considers scientific instruction as just one feature of a broader concept of education as edification.

However, in his interesting endeavor, Rorty faces us with unnecessary dichotomies. I have introduced and criticized two dichotomies; one putting sentiment against discussion and the other opposes objectivity in favor of solidarity. So far as the first dichotomy is concerned, it seems that there is no way but to combine sentiment and argument or else to embrace arbitrariness in dealing with different frames of reference. Neimann (1993) raises a similar critique to Rorty's dichotomy of irony and socialization by introducing the notion of 'ironic schooling' in which irony is intended to enter the primary schools, namely in the period of socialization according to Rorty.

As for the second dichotomy, it could not be denied that solidarity of society on attacking problems rests one way or another on objectivity. Having said all this, we cannot avoid admiring Rorty for his important criticism of representationalism because of its naïve interpretations in accounting for the relation between science and reality.

References

An Examination of Richard Rorty's Neo-Pragmatist Educational Views


CHAPTER SEVENTEEN

Reasons and Causes for an Aesthetic Response

Andrew Ward

First of all, we need to know what sorts of aesthetic response Wittgenstein has in mind. He is referring to such reactions as: ‘I am aesthetically dissatisfied with the height of that door in relation to its surroundings’; or to such question-and-answer examples as: ‘Why am I so dissatisfied with the facade of this building?’ - ‘Because the central door is too high’ - ‘Yes, you’re right, that’s what is dissatisfying me’. In other words, Wittgenstein is concerned with those aesthetic reactions that are expressive of a spectator’s aesthetic satisfaction or dissatisfaction with an object. Thus he says: ‘Perhaps the most important thing in connection with aesthetics is what may be called aesthetic reactions, e.g. discomfort, disgust, discontent. The expression of discontent says: ‘Make it higher...too low!...do something to this’ (Lecture II, Section 10).

On the face of it, these kinds of reaction do look as though they are making a causal claim. In the case of a spectator expressing aesthetic displeasure by saying, ‘The door in that facade is too high!’, his comments appear to be saying something like: ‘I feel dissatisfied with the design of this facade, and the cause, as I conjecture, is the door’s height’. Similarly, when you are puzzled as to why an object is aesthetically satisfying or dissatisfying you, it would appear that if you admit that a critic has put his finger on the specific reason for your puzzlement, you are supposing that he has, at least in part, located the cause of your feeling. Moreover, it would not seem absurd to hold that one central motive for listening to critics is to help us to locate our reasons for being satisfied or dissatisfied with aesthetic objects. Admittedly, it might well be held that critics have another, and even more significant, role, viz. showing us what we ought to be aesthetically satisfied or dissatisfied by, not what we are in fact moved by. However, as I shall argue later, these two roles are not as disparate as they may initially appear. For the moment, I wish to pursue not the question of value (‘What ought we to be moved by?’), but the question of fact (‘What are we moved by?’).

Now many able philosophers have, indeed, held that such critical procedures - whether factual or evaluative - do make a causal claim. Thus Hume, in his well known essay ‘Of the Standard of Taste’, maintains that those features in an aesthetic object that we term its ‘beauties’ or ‘deformities’ are so denominated because, or partly because, they are believed to be the cause of a true critic’s feelings of (respectively) pleasure or displeasure. He notes that were we to discover that what we now call the ‘faults’ or ‘deformities’ of a
poem were really the features that were producing the critic’s pleasure at the poem, we would need to change our beliefs about which features were its beauties and which its deformities: ‘Did our pleasure really arise from those parts of his [Ariosto’s] poem, which we denominate faults, this would be no objection to criticism in general: It would only be an objection to those particular rules of criticism, which would establish such circumstances to be faults, and would represent them as universally blameable. If they are found to please, they cannot be faults; let the pleasure, which they produce, be ever so unexpected and unaccountable’.

Yet, Wittgenstein sets his face against supposing that our expressions of aesthetic satisfaction or dissatisfaction can be seen as essentially making a causal claim or requesting to know the cause of our feelings: ‘I wish to make it clear that the important questions in aesthetics are not to be solved by psychological experiment’ (Lecture II, Section 36). And he stresses that the attempt by experimental psychologists to locate, in aesthetic objects, the causes of people’s pleasure or displeasure is simply irrelevant to a genuinely aesthetic or critical enquiry. But why did Wittgenstein seek to deny that our aesthetic reactions carry a causal implication?

To answer this question we need to understand how, as Wittgenstein sees it, the concept of cause figures in scientific investigations. Essentially, it is the same concept as Hume identified. There are two conditions that must jointly be satisfied for its application. First, the supposed cause (A) and effect (B) must be separately identifiable, otherwise A cannot be said to ‘bring about’ or ‘cause’ B in the scientific sense. Second, if A really is the cause of B, then A must be constantly conjoined with B under relevantly similar circumstances. Hence, if a case was discovered where, under these circumstances, A occurred but B did not follow, we should conclude that A cannot be the cause of B in any instance where these same circumstances obtain. Accordingly, the second condition for making a causal judgment is often put by saying that such a judgment must be vulnerable to parallel negative instances.

As I understand him, Wittgenstein holds that both these conditions for making a genuine causal judgment, viz. separate identification, on the one hand, and constant conjunction (and so vulnerability to parallel negative instances), on the other, are inapplicable when we are concerned with giving a critical reason for a spectator’s aesthetic reaction. If he is right, it must follow that critical reasons cannot be making causal claims, at least in Hume’s sense of cause (henceforth, reference to this qualification will be dropped but always implied).

II

Let us start with the separate identification condition. Wittgenstein contrasts somebody expressing aesthetic discomfort by saying ‘That door is too high!’

with somebody expressing bodily discomfort, e.g. stomach-ache, by saying, ‘I think I had too many tomatoes, today!’ In the case of the bodily discomfort, Wittgenstein accepts that there really are two separable things going on: there is the (present) stomach-ache and there is the (earlier) eating of the tomatoes. And he also acknowledges that the claim that it was the eating of the tomatoes that was responsible for, or brought about, the stomach-ache can be seen as a genuine causal hypothesis. For there is no necessity for the person to mention what he believes to be the cause of his stomach ache in order to express his particular bodily discomfort. He could, instead, simply have said in a pained way (perhaps while rubbing his stomach), ‘I have an awful stomach-ache!’. In short, with the expression of bodily discomfort, Wittgenstein concedes that the alleged cause and effect do meet the separate identification condition.

Moreover, the claim that it was the eating of the tomatoes that caused the person’s stomach-ache is a perfectly respectable scientific hypothesis in regard to the second condition for making a genuine causal claim, viz. vulnerability to parallel negative instances. If we can find another occasion where that person ate a similar number of tomatoes (and this other occasion is the same in all relevant respects) and, yet, he did not then experience stomach-ache, this will justify us in claiming that the tomatoes were not responsible for the stomach-ache, even on the original occasion. And however much the person in question may protest that it was the tomatoes that caused his bodily discomfort on the original occasion, the fact that this hypothesis has been subject to a genuine parallel negative instance shows that the person is mistaken. He is the final authority on whether he has an ache; but he is not the final authority on what was responsible for, what brought about, his discomfort.

Wittgenstein contrasts this, bodily discomfort, case with a person expressing aesthetic discomfort by saying, ‘That door is too high!’ Note the great verbal similarity between this latter expression of discomfort and the expression of bodily discomfort in the tomatoes example. Yet, Wittgenstein wants to say, this similarity masks what he calls a ‘grammatical’ difference - or, in fact, two such differences.

On the face of it, the most obvious difference is this. In the aesthetic case, there do not seem to be two separable events: first, aesthetic discomfort; second, a hypothesis as to the cause of this reaction, namely, the height of the door. Rather - as Wittgenstein himself emphasises - the very expression of discomfort, insofar as it is a genuinely aesthetic one, must say something about the object reacted to. An aesthetic reaction is necessarily directed onto features of the object, in the sense that there must be something about the object, however unspecific, of which the person is aware that he is satisfied or, as in this instance, dissatisfied. That is why Wittgenstein says: ‘The expression of [aesthetic] discomfort takes the form of a criticism, and not “my mind is not at rest” or something. It might take the form of looking at a picture and saying “What’s wrong with this?”’ (Lecture II, Section 19).

So, in the case of a genuine aesthetic feeling, Wittgenstein contends that a person must already be aware of what is satisfying or dissatisfying him. It is no good his looking at e.g. an Impressionist picture and simply saying: ‘Wow! I’m
getting a lovely tingling feeling down my spine!’. That, by itself, does not express an aesthetic satisfaction (this is not, of course, to deny that the person is experiencing a pleasurable sensation). For an expression of feeling to count as a genuinely aesthetic one, the spectator must be able to identify what is moving him in the very articulation of his feeling. In other words, his feeling must have an object that is identified, however unspecifically, in the actual expression of the feeling. Consequently, with the aesthetic case, there would not appear to be two distinct events, as there clearly are with the bodily discomfort case: it would seem that there is not the expression of aesthetic reaction, on the one hand, and the spectator’s awareness of its ground, on the other. As Wittgenstein insists, with respect to aesthetic discomfort, ‘To say “I feel discomfort and know the cause”, is entirely misleading because…[this] makes it sound as though there were two things going on in my soul – discomfort and knowing the cause’ (Lecture II, Section 16). On the contrary, the spectator can only justify his feeling as an aesthetic one, if its very expression mentions, at least in an unspecific way, what the feeling is directed onto. Contrast this with the bodily discomfort case, where, as we have seen, a person’s expression of a stomach-ache need mention nothing whatever about what is discomforting him in order to count as a genuine utterance of stomach-ache.

Still, as it stands, this position hardly represents a very forceful reply to the causal theorist. For, as Wittgenstein concedes, it can often happen that we are only able, anyway at the outset, to identify the object of our aesthetic feeling very unspecifically. Thus, we might listen to a piece of music and say ‘What’s wrong with this?’. Wittgenstein himself acknowledges that this would still be a case where we are expressing a directed response to an aesthetic object. Yet, ex hypothesi, we do not know what the precise ground of our response is: that is, indeed, just what we want to identify. Consequently, Wittgenstein’s position now looks open to the following objection: ‘Surely what we are asking for here are the specific features in the aesthetic object that are causing our dissatisfaction. So the fact that, in those familiar cases where we are puzzled about what precisely in the object is aesthetically moving us, the expression of our aesthetic reaction must already mention, albeit unspecifically, the object of our feeling does not rule out a causal analysis. It does not, since the alleged specific cause (particular features in the object) and the alleged effect (our aesthetic satisfaction or dissatisfaction) are here separately identifiable. They are separately identifiable because we have expressed our aesthetic feeling about an object without mentioning the precise features in the object that our feeling is directed onto’.

But Wittgenstein seeks to repudiate this objection on the ground that, in the aesthetic case, the spectator is always the final authority on what he is precisely moved by. He takes the following type of example: a spectator says, ‘There’s something wrong with the proportions of the side of this building. What precisely is it that is worrying me?’; and somebody suggests that the reason is that the door is too high. Now, Wittgenstein claims, this suggested explanation constitutes the precise ground of the spectator’s dissatisfaction if, but only if,
the spectator himself agrees that the explanation is the right one. In other words, the spectator’s own willing acceptance of any proffered reason for his directed feeling is the sole criterion of the correctness of that explanation. This, I take it, is what Wittgenstein is driving at in Lecture III, Section 10: ‘Suppose someone heard syncopated music of Brahms played and asked: “What is the queer rhythm that makes me [it?] wobble?”’. [Suggestion] “It is the 3 against 4”. One could play certain phrases and he would say: “Yes. It’s this peculiar rhythm I meant”. On the other hand, if he didn’t agree, this wouldn’t be the explanation.

If Wittgenstein is right that, in the case of aesthetic explanations, the spectator is always the final authority on the specific grounds for his feeling (and the contention is, I suggest, very plausible), then this does mark a real difference with many normal causal cases, like e.g. the bodily discomfort case. In that latter case, we noted that there is no requirement to gain the acceptance of the person with stomach-ache for the explanation - the eating of a certain number of tomatoes - to be the correct one; since, here, the person is not in any privileged position with regard to deciding whether it was his eating of the tomatoes that identifies the reason for his ache. Whereas, in the aesthetic case, Wittgenstein’s contention is that what constitutes the correctness of any proffered explanation is the spectator’s own willing acceptance that the explanation is the right one. For what the spectator wants is a description (or other mode of identification) of features of the aesthetic object which he can recognize as the precise ground of his directed feeling: a way of looking at the object which captures for him the ground of his delight or dissatisfaction (though he could not himself put his finger on that ground). So, in the aesthetic case, the spectator is in a privileged position with regard to deciding whether any proffered explanation identifies the precise reason for his reaction. From this difference, Wittgenstein concludes that aesthetic explanations are not causal explanations.

But why does Wittgenstein suppose that a difference of this nature shows that aesthetic explanations are not causal ones? His answer hinges on the idea that where we are the final authority on the ground of our response (as with an aesthetic reaction), it cannot be the case that we are implying a constant conjunction between that ground and our response. On the other hand, in judging that something is the cause of a certain phenomenon, we are implying

---

1To hold that the spectator is, in the sense outlined, the final authority on the precise ground of his directed feeling is not at odds with Wittgenstein’s own denial that there can be a logically private language of sensations. For there are plainly public criteria for determining whether, in any given case, a spectator is satisfied (or dissatisfied) with an aesthetic object, and there are plainly public criteria for determining whether a spectator is (or is not) conscious of those features of the object that he acknowledges as the precise ground for his directed feeling. Wittgenstein’s claim in Lectures on Aesthetics is that, given these public criteria can be met, we allow that the spectator’s own acceptance of certain features of the aesthetic object as the precise ground for his feeling as sufficient for them to constitute the precise ground. To maintain that a spectator is, in the sense explained, the final authority on what he is really moved by in an aesthetic object is no more at odds with that argument than it would be to maintain that a single man or woman is the final authority on what colour curtains each of them really wants in their own home.
that the assumed cause is constantly conjoined with that phenomenon (under the same circumstances). For, as we saw earlier, unless there is this constant conjunction, there cannot be a genuine causal relationship between the two.

Take the bodily discomfort case. If it is true that eating x number of tomatoes is the cause of a particular person’s stomach-ache, it follows that whenever a person of the same constitution eats x number of tomatoes (under relevantly similar circumstances), then stomach-ache will ensue. Furthermore, we would naturally seek to establish a causal connection here by showing that there is a constant conjunction between people of the same constitution eating that number of tomatoes and their having stomach ache. That is why Wittgenstein says about psychological experiments (which he acknowledges to be genuine causal hypotheses): ‘One of the curious things about psychological experiments is that they have to be made on a number of subjects. It is the agreements of Smith, Jones and Robinson which allows you to give an explanation - in this sense of explanation, e.g. you can try out a piece of music in a psychological laboratory and get the result that the music acts in such and such a way’ (Lecture III, Section 11).

Turn now to a request for an aesthetic explanation: for instance, where a person is unhappy about the aesthetic effect of a building’s facade and wants to know the precise reason for his dissatisfaction. If, as Wittgenstein maintains, that person is the final authority on what exactly is dissatisfying him, and he can always come to know the specific object of his dissatisfaction (the height of the door) without carrying out any experiments to establish a constant conjunction between, on the one hand, his (and perhaps others’) awareness of the height of doors, in relation to the facade of buildings, and, on the other, his (and perhaps others’) feelings of dissatisfaction then, it would seem, the type of explanation that we are looking for in aesthetics cannot be a causal one. For – the argument goes – we typically establish a causal claim by discovering the existence of a relevant constant conjunction while, in the case of an aesthetic explanation, the spectator can know the ground of his response independently of establishing whether there is a constant conjunction between that ground and his response.

III

I believe that Wittgenstein has drawn attention to some significant differences between aesthetic explanations and many typical causal ones. At the same time, I am doubtful that his arguments succeed in showing that aesthetic explanations are not a species of causal explanation. For, as we have seen and as Wittgenstein himself acknowledges, at least in those cases where the spectator is puzzled as to the specific ground for his directed feeling, that feeling can be expressed separately from his identification of its precise ground (thereby explaining how he can feel satisfied or dissatisfied with an aesthetic object while still wondering precisely why). Hence, the first condition for making a genuine causal judgment, viz. separate identification, is met, at least
in the cases where a spectator expresses his aesthetic reaction without identifying its precise ground. Moreover - and this seems to be a factor that Wittgenstein does not consider - we think of any acceptance by the spectator of certain features as the precise ground of his aesthetic reaction as subject to a familiar form of causal falsification, viz. where we point to a parallel negative instance in which those same features were identified by him as present while the feeling was absent. Thus, the second condition for making a causal judgment, viz. vulnerability to parallel negative instances, is also met in these cases. For example, we question whether, on a given occasion, certain features can be the real reason for a spectator’s dissatisfaction with an aesthetic object - even though he accepted them as the precise ground - if, on another similar occasion, their identified presence did not dissatisfy him. And this doubt will be reinforced if we can discover another set of features in the present situation that has regularly dissatisfied him on earlier occasions.

Let us see how these two conditions, separate identification and vulnerability to a parallel negative instance, would typically operate together in a particular case of aesthetic puzzlement. Say that a spectator expresses himself dissatisfied with the appearance of a cathedral vault, but is puzzled as to what precisely is dissatisfying him. And let us say that following the suggestion from another, he accepts as the specific reason for his dissatisfaction ‘the star-shaped design of the crossing ribs’. But suppose that there have been earlier occasions where the same, or a very similar, rib design had been identified by him on other church vaults, and he had felt no dissatisfaction at this design. Suppose, too, that on this present occasion the bosses on the vault were highly coloured, and that he had always expressed dissatisfaction, on earlier occasions, at this sort of colouring of bosses (though there had never been a past case where he had experienced vaults of the present rib design with coloured bosses). In this kind of situation, we say things like: ‘The feature that he accepted as the reason for his dissatisfaction - ‘the star-shaped rib design’ - cannot really be the precise ground for his dissatisfaction. The real reason probably arises from the colouring of the roof bosses’. In other words, we proceed here in the same way as we would in any indisputably causal claim, where the production of a parallel negative instance (perhaps together with the discovery of some other factor in the circumstances which, on other occasions, has been regularly accompanied with the same end result), leads us to say that the proffered ground cannot, after all, have been the real reason why the given event occurred (and that what was probably responsible for it was the presence of the other discovered factor).

Admittedly, even a fully rational spectator may not always be prepared to back down, anyway initially, in the face of a claimed parallel negative instance. For the spectator may honestly believe that his aesthetic taste has changed since the occurrence of the past negative instance (so that what looks like a parallel negative instance is not, in his opinion, genuinely one). But he cannot simply refuse to admit the relevance of an apparently parallel negative instance, or – crucially – reasonably continue to claim that his taste has changed if he discovers that, after all, that earlier instance is still satisfying
him. In this respect, there is no difference with indisputably causal judgments. When an apparently past parallel negative instance is produced to show someone that his recent stomach ache was not, as he claimed, the result of eating too many tomatoes (and we suspect that it resulted from some other ingredient in the dish which is frequently found to produce stomach ache), the person may seek to repudiate that proposed past parallel negative instance by contending that, since the past occasion of eating the same number of tomatoes, his constitution has become more delicate, i.e. has changed in the relevant respect. But he cannot reasonably continue to maintain this, if it is now found that his eating a further relevantly similar tomato dish to the recent one (except that it lacked the ingredient that we suspect did produce his stomach ache on that recent occasion) has no adverse effects. In short, it can no more be an objection to the claimed causal implication of aesthetic explanations that the spectator may, in certain cases, have changed his taste, since the occurrence of an earlier apparent counter instance, than it can be an objection to indisputably causal explanations that they can, in certain cases, be over-ruled because of a change in the circumstances of the effected agent, since the occurrence of an earlier apparent counter instance. I conclude that in accepting certain features, under a given description, as the precise reason of his aesthetic reaction, a spectator is hypothesising those features, so identified, as the cause, or an important part of the cause, of his response.

The upshot is that although a spectator may be the final authority concerning the precise reason for his directed feeling in one sense - in the sense that he may be able to rule out suggestions from others (so that unless he willingly accepts the proffered features, under the given description, as at least part of the precise reason for his reaction, they cannot, so identified, be part of that reason) - there is another sense in which the spectator’s say-so is not overriding. For the spectator cannot justifiably continue to affirm that these features, as so described, are the precise reason for his reaction, if it can be shown that they are not a significant part of its cause, i.e. by the production of a parallel negative instance.

Consequently, despite some notable differences with typical causal explanations, aesthetic explanations (of the kind Wittgenstein is considering) do make a causal claim. Agreed, the spectator’s ground for asserting that e.g. it is the door’s height that is dissatisfying him need not be based on experiments - on consciously noticing a past constant conjunction between, on the one hand, certain door heights in relation to the facades and, on the other, ensuing dissatisfaction - but it does imply that the two are constantly conjoined, however that claim may have been originally arrived at. That is why we are prepared to back down in the face of parallel negative instances, with remarks like: ‘It can’t, after all, be the height of the door that is really dissatisfying me’. And this consideration is equally applicable in those cases where the spectator has expressed his aesthetic reaction by actually mentioning what he takes to be the precise ground of his feeling: it cannot simply be confined to those cases where the spectator is unable himself to identify what he is precisely moved by in the object, and seeks help from others. For if there is a causal implication in
the latter cases, there must be one in the former cases also. After all, in both cases, the spectator’s aesthetic reaction can be expressed unspecifically, i.e. by simply saying that he is moved by something in the aesthetic object. Therefore, the fact that, in some cases but not in others, the spectator has identified what he takes to be the precise ground of his reaction cannot be a reason for denying that in such cases his expression of aesthetic reaction carries a causal implication, given that his reaction can always be re-expressed without mentioning what he believes its precise ground to be.

The sum of what I have so far argued is this. I have acknowledged that features of an object, under a given description, can only be the precise reason for a spectator’s aesthetic reaction, if he is ready to accept that, as so identified, they are the reason for his response. His acceptance - laying aside problems of self-deception - functions as a necessary condition for the discovery of the precise ground. But I have also argued that the spectator’s acceptance is not a sufficient condition. It is further required that these features, so identified, must also be an important part of the cause of his reaction. Hence, although Wittgenstein has pinpointed some real differences between expressions of aesthetic feeling and certain other expressions of feeling which do clearly have a causal implication, he has not succeeded in disproving Hume’s contention that in offering reasons for our aesthetic reactions, we are implying that these reasons function as a significant part of the cause of our reactions.

IV

How is all this related to the question of aesthetic value? Granting that what the critic principally aims to tell us is what we ought to be moved by, how can our merely knowing what is in fact the precise ground of our aesthetic reaction be relevant? Of course, in itself, it cannot be. At the same time, it does seem right to hold that, in seeking to appraise an aesthetic object, one of the critic’s central tasks is the identification of those features in the object that are the precise reason for his own aesthetic reaction to it. And on the assumption that the critic is what Hume calls a ‘true critic’, it follows that in locating the precise reason for his own satisfaction or dissatisfaction, he must be locating the specific features, or some of the specific features, in that aesthetic object which ought, correspondingly, to ground any spectator’s aesthetic pleasure or displeasure. For a ‘true critic’ is someone who, in studying an aesthetic object, is able both to feel very fully its effects upon him, and, correspondingly, to identify those features which are, as he supposes, the precise ground of his reaction. Indeed, it is just the true critic’s identifications that initially determine what are to count as the aesthetic features, the beauties and the deformities, of an object - and, thereby, the precise reasons for a justifiable aesthetic reaction.

However, if it were discovered that these features, as he identified them, were not, after all, the cause of the true critic’s own response, then they must cease to be regarded as the beauties or deformities of the object. They must cease to be so regarded since, as we have seen, it is necessary - though not
sufficient - that they be an important part of the cause of the true critic’s directed feelings of, respectively, pleasure or displeasure.

It transpires, then, that Hume’s claim that an object’s beauties or deformities must be the cause of a true critic’s pleasure or displeasure can be defended against Wittgenstein’s objection that aesthetic (directed) reactions do not carry, or importantly carry, a causal implication. Contrary to this objection, I have argued that the true critic’s identification of certain features as the specific ground of his aesthetic reaction is partly a causal hypothesis. In fact, the very conception of what is aesthetically valuable depends on this causal implication. That is why, as Hume notes, we should alter our view of what features count as the beauties or faults of an object, if those features, so denominated, were found not to be causally efficacious in the way that we had supposed.

Nonetheless, it would be a serious mistake to contend that locating those features of objects that are the causes of our aesthetic reactions will be enough to solve aesthetic puzzlements. As Wittgenstein has shown, what is also required for their solution is a description (or other mode of expression) that the spectator can himself recognize as a fitting one. It is unclear to what extent Hume, and contemporary thinkers who support the Humean position, have grasped the significance of discovering descriptions of the object that can satisfy the spectator. At any event, such illuminating identifications are, I submit, of central importance to the solution of our aesthetic puzzlements and to aesthetic criticism generally. They are of central importance because what a spectator is searching for is a way of thinking about the causes in the object that can bring home to him why he is appreciating or ought to be appreciating that object. This requires considerably more than discovering the features in the object that are causing his feelings. For by no means every way of identifying these causes will be ones that the spectator can recognize as answering his puzzlement about the object. In the matter of aesthetic appreciation, finding the right mode of describing the cause – viz. the one that satisfies the spectator – is, I submit, as important as finding the cause itself.

Although, as I have argued, Wittgenstein was wrong to deny Hume’s claim that aesthetic explanations are a species of causal explanation, he has, in the process of this very denial, succeeded in bringing out a vital feature of aesthetic appreciation, and one that is all too easily overlooked by advocates of the causal thesis. Let us by all means acknowledge that the precise reasons for our aesthetic feelings must be their causes; but let us not forget that it is how these causes are described or identified that truly enriches our aesthetic experience.
Part 2

Metaphysics, Epistemology and Philosophy of Language
Logically Unknowable Propositions:
A Criticism of Tennant’s Three-Partition of Anti-Cartesian Propositions

Massimiliano Carrara & Davide Fassio

The Knowability Paradox is a logical argument that, starting from the plainly innocent assumption that every true proposition is knowable, reaches the strong conclusion that every true proposition is known. The same conclusion could also be put as follows: if there are unknown truths, there are unknowable truths. The paradox has been considered a problem for every theory assuming the Knowability Principle, according to which all truths are knowable (formally: (KP) ∀q (q → ◊Kq)), and, in particular, for semantic anti-realist theories.

A well-known criticism of the Knowability Paradox is the so-called restriction strategy. This bounds the scope of the universal quantification in (KP) to a set of formulae, whose logical form avoids the paradoxical conclusion. Specifically, Tennant suggests restricting the quantifier in (KP) to propositions whose knowledge is provably consistent. He calls them Cartesian propositions. They are distinguished from Anti-Cartesian propositions, which are propositions whose knowledge is provably inconsistent, and which are responsible for the paradox. Tennant distinguished Anti-Cartesian propositions in three kinds. In this paper, we will not be concerned with the soundness of the restriction proposal. Rather, we are interested in analyzing the proposed distinction.

We argue that Tennant’s distinction is problematic because it is not grounded on an adequate, logical analysis, and because it is incomplete. We suggest an alternative distinction, and give the following reasons for accepting it: it is logically grounded and more complete than those of Tennant’s one; are inclusive of Tennant’s distinction; and independent from non-epistemic notions.
The Knowability Paradox

The Knowability Paradox is a logical argument published by F. Fitch in an article entitled A Logical Analysis of Some Value Concepts in 1963. It runs as follows.

Adopt a standard system of modal logic with an epistemic operator K, where Kp stands for “someone knows that p”; “it is known that p”, and p is a proposition.2

Assume the validity of two properties of knowledge: 1) distributivity over conjunction (if a conjunction is known, then so are its conjuncts), and 2) factivity (if a proposition is known, then it is true). Formally:

\[(\text{Dist}) \ K(p \& q) \vdash Kp \& Kq\]
\[(\text{Fact}) \ Kp \vdash p\]

Assume the following common modal inferences:

\[(\text{Nec}) \text{ if } \vdash p, \text{ then } \Box p\]
\[(\text{ER}) \ \Box \neg p \vdash \neg \Diamond p\]

(Nec) states that every theorem is necessary, and (ER) states that if it is necessarily false that p, then it is impossible that p. Assume also the Knowability Principle:

\[(\text{KP}) \forall q (q \rightarrow \Diamond Kq)\]

Every true proposition is knowable. Finally, assume that we are not omniscient, i.e. there is at least a truth that is not known, and instantiate an example of it; formally:

\[(2) \ p \& \neg Kp\]

Read “p and it is not known that p”. Consider an example of (KP) resulting from the substitution of q in (KP) with (2):

\[(3) \ p \& \neg Kp \rightarrow \Diamond K(p \& \neg Kp)\]

By (2) and (3), by Modus Ponens, we obtain:

\[(4) \Diamond K(p \& \neg Kp)\]

---

1Fitch (1963). For a detailed account of the argument and of the literature about it, see Brogaard & Salerno (2002), and Kvanvig (2006).

2Some further remarks on the language: the letters ‘p’ and ‘q’ are arbitrary propositional letters (letters for propositions). You can read the symbols & , ¬, 0, □, and ∀ as “and”, “it is false that”, “it is possible that”, “it is necessary that”, “for every” respectively. “├” is the symbol of logical derivability: p ├ q means that q is derivable from p.
Consider the following argument *per absurdum* (independent from (2)-(4)):

(5) K(p & ¬Kp)  
(6) Kp & K¬Kp  
(7) Kp & ¬Kp  
(8) ¬K(p & ¬Kp)  
(9) □¬K(p & ¬Kp)  
(10) ¬◊K(p & ¬Kp)

(by (5) and (Dist))

(by (5)-(7), negating (5))

(by (8) and (Nec))

(by (9) and (ER))

(4) is inconsistent with (10). If so, from (2) and (KP), an inconsistency is derived. The two premises are incompatible.

There are (at least) two strategies one might adopt in order to avoid the paradox: one could either negate (2) or (KP). If one negates (2), p & ¬Kp, that is equivalent to:

(11) p → Kp

If p is true, it is known. So, accepting (KP) and discharging (2), if every truth is knowable ((KP)), then every truth is also actually known (from (11) by generalization). But, the last thesis is clearly indefensible. So, in order to avoid the contradiction, the other assumption, (KP), must be negated, i.e.:

(12) ¬∀p (p → ◊Kp)

informally: not every true proposition is knowable.

II. The Restriction Strategies and Tennant’s Anti-Cartesian Propositions

A seemingly weak demand of the so-called semantic anti-realism is that every truth must be knowable, i.e. the Knowability Principle (KP) is true. But, as we have seen in the previous section, according to the Knowability Paradox, if one accepts (KP), she must accept that all truths are known; a very demanding conclusion. So, the Knowability Paradox seems to be a difficult challenge for semantic anti-realism.

One way in which to reply is to deny that the anti-realist’s commitment to the epistemic character of truth involves any commitment to the Knowability Principle. Many philosophers have followed this train of thought, suggesting restrictions of the Knowability Principle that are able to bound the scope of universal quantification in (KP) to a set of formulae, whose logical form avoids the paradoxical conclusion. This way of solving the paradox is commonly called the restriction strategy.

---

1 (11) is classically but not intuitionistically deducible by the negation of (2). Some criticisms against the paradox have been carried on this line. See, for example, Williamson (1982).
One of the most famous and debated proposals is that of Neil Tennant in _The Taming of the True_ (Tennant (1997), pp. 272-276). His restriction is based on the distinction between Cartesian and Anti-Cartesian propositions. Propositions whose corresponding knowledge claims are consistent will be called Cartesian:

\[ \text{Kp} \nvDash \bot \]

(read: from Kp is not deducible an inconsistency) By contrast, every proposition whose knowledge is provably inconsistent is Anti-Cartesian:

\[ \text{Kp} \vdash \bot \]

Tennant proposed the distinction between Cartesian and Anti-Cartesian propositions to formulate a restricted principle, allowing the latter to escape from quantification. The restricted principle is as follows:

(KCP) \[ q \rightarrow \Box Kq, \text{where “q” is Cartesian} \]

His restriction avoids the paradox because of the second assumption in the argument, (2) p & ¬Kp, is an Anti-Cartesian proposition, since its knowledge is provably inconsistent (from the _reductio_ in the paradox, steps (5)-(7)). Therefore, we cannot substitute (2) for q in the restricted _Knowability Principle_ (KCP). The first step in the argument:

1. \[ (p & \neg \text{Kp}) \rightarrow \Box (p & \neg \text{Kp}) \]

is incorrect and the paradox is stopped at the beginning.

A common criticism of Tennant’s strategy is that it is _ad-hoc_. The criticism seems to be correct because there is not a clear reason for excluding Anti-Cartesian propositions from the quantification inside the _Knowability Principle_, except to escape the paradox. This solution seems obviously unprincipled. Tennant has tried to respond to this criticism. However, we are not concerned here with Tennant’s restriction and its problems. We are more interested in Tennant’s distinction of the Anti-Cartesian propositions into three different kinds.

According to Tennant,

---

1 Other well-known restrictions are those of M. Dummett (2001) and D. Edgington (1985). For a short list, see Kvanvig (2008).

2 J. Hintikka’s _epistemically indefensible_ propositions (Hintikka (1962b)), Routley’s _unknowable propositions_ (Routley (1981)), and Soerensen’s _epistemic blindspots_ (Soerensen (1988)) are characterizations similar to those of Tennant.

3 For this criticism see, for instance, Hand-Kvanvig (1999).

4 See, for example, Tennant (2001).
«[t]here are three broad kinds of Anti-Cartesian proposition $\phi$, corresponding to the kind of reason why knowledge that $\phi$ is impossible:

First, the proposition $\phi$ itself may be inconsistent; whence the proposition that $\phi$ is known will be inconsistent. So, for example, any compound proposition of the form ($\phi \& \neg\phi$) is Anti-Cartesian.

Secondly, knowledge of a (consistent) proposition $\phi$ may be impossible because the very act of considering or judging (falsely) that $\phi$ requires the falsity of (some consequence of) $\phi$. *A fortiori* the proposition that $\phi$ is known is inconsistent. It is in this way that the proposition that no thinkers exist is Anti-Cartesian.

Thirdly, the proposition that $\phi$ is known may be logically inconsistent because of its own overall logical structure, involving iterations of K (and perhaps of other attitudes). Thus for any $\phi$ the proposition ($\phi \& \neg K \phi$) is such that that ($\phi \& \neg K \phi$) is known turns out to be logically inconsistent [(as the Knowability Paradox shows)] [...] That is, ($\phi \& \neg K \phi$) is Anti-Cartesian» (Tennant (1997), pp 272-273).

Unfortunately, Tennant does not explain why he brings in these, and only these, kinds of propositions: he does not argue for his distinction, but simply lists it. We have found only some hints for justifying it.

In respect of the first kind of propositions, their inclusion in the list seems quite clear: every factive operator – as is the case with the epistemic one – applied to an inconsistency implies the truthfulness of such inconsistency. But inconsistencies are necessarily false.

The reason behind Tennant’s motivation to introduce the second kind of Anti-Cartesian propositions is particularly unclear. Some authors have argued that these propositions could be identified with the *existentially inconsistent* propositions of J. Hintikka (1962a)\(^1\). Hintikka defines *existentially inconsistent* propositions as follows:

«Let $p$ be a sentence and $a$ a singular term (e.g. a name, a pronoun, or a definite description). We shall say that $p$ is *existentially inconsistent* for the person referred to by $a$ to utter if and only if the longer sentence ―$p$; and $a$ exists‖ is inconsistent. [...] Uttering such a sentence [...] means making a statement which, if true, entails that its maker does not exist» (Hintikka (1962b), p. 11).

Tennant’s example “no thinkers exist” could be considered a case of existential inconsistency and this seems to be the reason for its unknowability. Tennant seems to move away from every particular interpretation of the origin of the term “Cartesian” (he writes: «"Cartesian” has been chosen for convenience» (Tennant (1997) p. 273, footnote 25). However, given that the title of Hintikka’s article is just “Cogito, Ergo Sum: Inference or

\(^1\)For M. Hand and J. Kvanvig, «[Anti-Cartesian proposition] might be the sort of proposition Descartes considered, such as no thinking thing exists. These propositions are consistent but existentially inconsistent [...] The idea of existential inconsistency is due to Jaakko Hintikka (1962a)» (Hand & Kvanvig (1999), p. 423).
Performance?”, the term “Cartesian” could be considered a clue to the above interpretation.

In respect of the third kind of Anti-Cartesian propositions, the Anti-Cartesianity of the logical form \( p & \neg Kp \) has been deduced by the Knowability Paradox; Tennant includes this form in a range of propositions whose knowledge is inconsistent as a consequence of some iteration of \( K \), but he does not point out any other clarifying examples.

Given the above considerations, it seems that Tennant’s distinction is not particularly accurate and it is not grounded on any sort of analysis. Moreover, it can be the target of criticisms pointing out its incompleteness. Take, for example, the following case: imagine there is a machine, such that if one pushes on a button, the machine affects his memory and he forgets having pushed on the button.\(^1\) The proposition “I pushed on the button” is Anti-Cartesian. In fact, if known, it is true (by (Fact)); but if true, it implies that it is unknown, because it has been forgotten. Therefore, if known, it is unknown, which results in an inconsistency. But the proposition is not, in itself, inconsistent (kind 1), nor is it a case of knowledge iteration (2) or of existential inconsistency (3). The proposition seems not to be considered in Tennant’s distinction.

In conclusion, Tennant’s distinction seems to be not well grounded and incomplete.

### III. Analysis of Anti-Cartesian propositions

In this section of the paper, we give an analysis that is able to settle what logical forms should be derivable by a proposition, for this proposition to be Anti-Cartesian. Then, in the next section, we will use the results of the analysis to give a new foundation of Anti-Cartesian propositions’ distinction, able to settle the faults of Tennant’s distinction.

Let us start with the analysis. We will proceed as follows:

1. We take an operator \( K \), where the only properties of \( K \) are distributivity and factivity.\(^2\)
2. We assume \( Kp \) (where \( p \) is an arbitrary Cartesian proposition) and one of the two properties mentioned above (for example, (Fact)).
3. We list all the propositions deductible in propositional logic by those assumptions.

---

\(^1\)The example is freely adapted from Egrè (undated).

\(^2\)A complete analysis of Anti-Cartesian propositions should consider all the logical properties owned by \( K \), but here, we will be concerned only with the above properties because otherwise, the analysis would require much more space. However, we think that our analysis could be considered, if not complete, at least, very reliable, because Anti-Cartesianity seems to grow from no more than these two properties. So, we could suspect that these properties and no others are responsible for such a phenomenon. Our intention is to extend the future the analysis to the complete axiomatization of some specific epistemic logic.
Logically Unknowable Propositions:
A Criticism of Tennant’s Three-Partition of Anti-Cartesian Propositions

(4) The key passage of the analysis is: if p denies at least one proposition deduced from the assumptions, then from its knowledge an inconsistency is deducible, and it is therefore Anti-Cartesian. On the contrary, if from p there is no deducible negation of at least a proposition deduced by the assumptions, then p is not Anti-Cartesian, because no inconsistency can be derived. In this way, we individuate what logical forms should be derivable by a proposition for this proposition to be Anti-Cartesian.

(5) We repeat the procedure with the other property of K and with the two properties taken together.

(6) Recursively, we give an analysis of complex propositions, including the ones obtained in the previous steps of the analysis. Given this analysis, we obtain the propositions that should be derivable from p, for p to be Anti-Cartesian.

Here are the details of the analysis:

Case 1. Consider just (Fact); proposition p, such that \( p \vdash \neg Kp \) and \( p \vdash \neg p \) is Anti-Cartesian.

Proof. Assume Kp. From (Fact), Kp → p, in propositional logic, by the common rules of inference, we can only deduce p, Kp (the assumption) or a disjunction including those propositions. From those premises, the sub-cases to consider are: 1.1) Kp, 1.2) p, 1.3) p & Kp (the conjunction of case 1.1 and 1.2), 1.4) every disjunction including one of the above cases. Consequently, p is Anti-Cartesian,

(1.1) if \( \neg Kp \) is derivable from p. Because, given the assumption Kp, we derive (Kp & \( \neg Kp \)), an inconsistency;
(1.2) if \( \neg p \) is derivable from p. Because we can deduce p from (Fact) and the assumed Kp, and we derive (p & \( \neg p \)), an inconsistency;
(1.3) if \( \neg (p & Kp) \) is derivable from p. Because it would contradict the conjunction of the two cases listed above (1.1 and 1.2);
(1.4) if the negation of a disjunction, including one of the above propositions, is derivable from p. In fact, the negation of a disjunction is the negation of each disjunct. But if from p is deducible the negation of one of the above propositions in conjunction with another proposition, from p is deducible the negation of one of the above propositions alone, for A & B \( \vdash A \).

Some remarks. In the first case, we have a proposition p from which is deducible its own ignorance; p is such that: \( p \vdash \neg Kp \). In the second case, p implies its own negation, \( p \vdash \neg p \), i.e. it is self-contradictory. Notice that this second kind of proposition includes inconsistencies (necessarily false propositions).
Case 2. From (Dist) alone, by the application of K to an arbitrary conjunction (say (p & q)), an inconsistency cannot be deduced.

Proof. Assume K(p & q), where p & q is whatever conjunction (the same proof could be repeated with more than two conjuncts). From those premises, we can only deduce Kp, Kq, Kp & Kq, or whatever disjunction, including one of those propositions. So, from those premises, the possible cases are: 2.1) Kp, 2.2) Kq, 2.3) Kp & Kq and 2.4) every disjunction, including one of the above propositions.

We have that, assuming (Dist) alone, p or q (or both) cannot be Anti-Cartesian. In fact, note that whichever is the logical form of propositions p and q, an inconsistency cannot be deduced by (Dist) alone, because distributing K over propositions, whatever negation in p and q remains under the scope of K, we cannot have ¬Kp or ¬Kq. But, for contradicting Kp, Kq or their conjunction (cases 2.1-2.3), we request ¬(Kp & Kq), that is ¬Kp v ¬Kq, where the negation is outside the scope of the operator. So, by (Dist) alone, and cases 2.1-2.3, there are not Anti-Cartesian propositions.

Case 3. From (Dist) and (Fact) together, propositions with logical forms p & q, where q ├ ¬p, or p ├ ¬q, or q ├ ¬Kp, or p ├ ¬Kq are Anti-Cartesian.

Proof. Assume K(p & q), where p & q is whatever conjunction (the same proof could be repeated with more than two conjuncts). From those we can only deduce p, q by (Fact), Kp, Kq by (Dist), a conjunction of two or more of those propositions, and whatever disjunction, including one of those. From those premises, the possible cases are: 3.1) p, 3.2) q, 3.3) Kp, 3.4) Kq, 3.5) p & q, 3.6) Kp & Kq, 3.7) Kp & p, (Kq & q), 3.8) Kp & q, (Kq & p), 3.9) longer conjunctions (cases explainable appealing to 3.5-3.8), 3.10) every disjunction, including one of the above propositions (3.1-3.9).

We have that, p or q (or both) are Anti-Cartesian,

(3.1) a) if ¬p is derivable from p. See 1.2.
   b) if ¬p is derivable from q. In this case, we have that p & q is inconsistent, entailing p & ¬p.
(3.2) see 3.1 inverting the cases ((a) for (b) and vice versa).
(3.3) a) if ¬Kp is derivable from p. See case 1.1.
   b) if ¬Kp is derivable from q. Because in this case, we would have Kp (by Dist) and ¬Kp (from q, by Fact).
(3.4) see 3.3 inverting the cases.
(3.5) a) if ¬(p & q) is derivable from p. That is ¬p or ¬q. The reason is shown taking together points 3.1 and 3.2.
   b) if ¬(p & q) is derivable from q. See 3.5a.
(3.6) a) if ¬(Kp & Kq) is derivable from p. That is ¬Kp or ¬Kq. See points 3.3 and 3.4 together.
   b) if ¬(Kp & Kq) is derivable from q. See 3.6a.

1Notice that here we assume just (Dist). The point here is that Anti-Cartesianity does not emerge by (Dist) alone.
Logically Unknowable Propositions: A Criticism of Tennant’s Three-Partition of Anti-Cartesian Propositions

(3.7 a) if \( \neg(Kp \& p) \) (\( \neg(Kq \& q) \)) is derivable from \( p \). That is \( \neg Kp \) or \( \neg p \) \( (\neg Kq \) or \( \neg q) \). See points 3.1a and 3.3a (3.1b, 3.3b) together.

b) if \( \neg(Kp \& p) \) (\( \neg(Kq \& q) \)) is derivable from \( q \). See 3.7a.

(3.8 a) if \( \neg(Kp \& q) \) is derivable from \( p \). That is \( \neg Kp \) or \( \neg q \). See points 3.1 - 3.4.

b) See 3.8a.

(3.9) conjunction of cases 3.1-3.8.

(3.10) if the negation of a disjunction, including one of the above propositions, is derivable from \( p \). See case 1.4

Some remarks. Observe that, in 3., the cases of Anti-Cartesianity not reducible to a sub-case of case 1 or to cases 3.1 - 3.4, are only the cases 3.1b) \( q \mid \neg p \), 3.2b) \( p \mid \neg q \), 3.3b) \( q \mid \neg Kp \), and 3.4b) \( p \mid \neg Kq \). The first two cases are Anti-Cartesian because the contradictory proposition of one of the two conjuncts is deducible from the other conjunct.

Notice that if \( p \& q \) is contradictory, (Fact) is sufficient to show its Anti-Cartesianity. In fact, as said in the explanation of case 1, case 1.2) \( p \mid \neg p \) includes all self-contradictory propositions, including inconsistencies (necessarily false propositions). On the contrary, cases 3.3b) and 3.4b) are not deducible by (Fact) or (Dist) separately, but only if taken together. The logical form of the propositions responsible of the Knowability Paradox, \( p \& \neg Kp \), is an example of those cases: it is obtained substituting \( \neg Kp \) for \( q \) in the conjunction \( p \& q \) (and, therefore, \( q \mid \neg Kp \), case 3.3b)

At this point we have exhausted all the possible cases where \( K \) is a distributive and factive operator and \( Kp \mid \bot \). We can consider propositions included in those cases as “basic” Anti-Cartesian propositions, but for the list to be complete, we need to show how those propositions behave if included in complex propositions. This job can be achieved recursively, showing if, how and in what cases the composition of “basic” Anti-Cartesian propositions results in “complex” Anti-Cartesian propositions.

Case 4. Complex cases.

4.(K) If \( p \) is Anti-Cartesian, then \( Kp \) is also Anti-Cartesian. In fact, given (Fact), if \( Kp \) is Anti-Cartesian, it results in an inconsistency (by the definition of Anti-Cartesian). But an inconsistency is, itself, an Anti-Cartesian proposition (see 3.1b, 3.2b, or simply 1.2).

4.(&) If \( p \) is Anti-Cartesian, then \( p \& q \), for every \( q \), is also Anti-Cartesian. In fact, if \( p \& q \) is known, from (Dist) we have that each conjunct is known (\( p \) included). But if \( p \) is Anti-Cartesian, from one’s knowledge an inconsistency follows. If, in a conjunction, one of the conjuncts is inconsistent, all the conjunction is inconsistent. So, knowing \( p \& q \) implies inconsistency, i.e. \( p \& q \) is Anti-Cartesian.

4.(v) The case of disjunction is a bit more complex. In particular, we have examples of known disjunctions of Anti-Cartesian propositions. Take the example of a proposition that we know to be either true or false, but we do not know whether it is true or false. In this case we have a proposition like: \( K((p \&
¬Kp) v (¬p & ¬K¬p)). As the Knowability Paradox shows, knowledge of both the disjuncts of the above proposition gives inconsistency, but their disjunction is knowable. The same discourse is valid for other Anti-Cartesian propositions.

The only exceptions are propositions like 1.2) p ├ ¬p. Those propositions are necessarily false because they are self-contradictory; and a disjunction of inconsistencies is, itself, an inconsistency; but, by (Fact), if a disjunction including inconsistencies is known, it is true, and this is contradictory. The same applies if we have a disjunction of Anti-Cartesian propositions, including propositions like (1.2), and only one Anti-Cartesian proposition deriving from a different case. In fact, knowing a disjunction of n disjuncts, where n-1 disjuncts are inconsistencies, is the same as knowing the only possibly true disjunct; but we assumed that it is Anti-Cartesian. So, knowledge of the whole disjunction gives an inconsistency, and, therefore, it is Anti-Cartesian.

At this point we have concluded our analysis, showing what logical forms should be derivable by a proposition for this proposition to be Anti-Cartesian. Every Anti-Cartesian proposition could be reduced to one of the shown cases. In fact, we have considered for each property and for their conjunction all the possible cases in which any proposition, if known, is inconsistent. By (Dist) and (Fact) we could deduce only the basic cases 1-3, and, recursively, only the complex sub-cases in case 4. So, our analysis can be considered complete.

A noteworthy finding is that the complex cases obtained recursively in case 4 and the other sub-cases in 1-3 are all reducible to three basic sub-cases: 1.1) p ├ ¬Kp, 1.2) p ├ ¬p, and 3.3b) q ├ ¬Kp (or equivalently 3.4b). Whatever Anti-Cartesianity is reducible to one of those basic cases. For example, every inconsistency can be reduced to 1.2, as in sentences like “no thinkers exist”\(^1\) (assuming that thinking is a necessary condition for knowing) to case 1.1, and Moorean sentences (“p and it is not believed that p”) to case 3.3b.\(^2\) So, it seems that every case of a logical unknowable proposition is (or is reducible to) a self-contradiction (1.2), or a proposition from which is derivable its own ignorance (1.1), or a conjunct from which is derivable the ignorance of the other conjunct (3.3b).

\(^{1}\)As Hand & Kvanvig (1999) noted, the unknowability of those propositions faces semantic Anti-Realism with a further issue known as “Idealism Problem”: if propositions like “no thinkers exist” are unknowable, and if knowability is a necessary condition of truth, then those propositions are necessarily false. The consequence is that it is impossible that no thinkers exist, that is, it is necessarily true that there are thinking creatures.

\(^{2}\)Notice that whatever necessary condition for knowledge exists, it can be substituted with the K operator in 1.1 and 3.3b cases, preserving Anti-Cartesianity. For example, admitting that the existence of subjects is a necessary condition for knowing, the proposition (*) “there are not subjects” is Anti-Cartesian. In fact, if there are no subjects, there are no known propositions, and the same proposition (*) is one of them. So, (*) is a proposition included in the case 1.1, from which is deducible its own ignorance. On the contrary, if we do not admit that the existence of subjects is a necessary condition for knowing, (*) is perfectly knowable and not Anti-Cartesian. An analogous reasoning is valid with every other necessary property of knowledge. “No propositions are believed” and “p and it is not believed that p” are other examples of this kind.
IV. A New Distinction of the Anti-Cartesian Propositions and a Comparison with Tennant’s Distinction

In section II, we have introduced Tennant’s distinction of Anti-Cartesian propositions. In section III we have proposed an analysis showing what logical forms should be derivable by a proposition for this proposition to be Anti-Cartesian. In this section, we are going to introduce a new distinction based on such analysis. Then, we will suggest some reasons for preferring our distinction, by showing some advantages deriving from it.

Our distinction is a bipartition between propositions derivable by (Fact) alone (case 1), call them kind I, and propositions derivable by (Dist) and (Fact) together (case 3); let us call them kind II. Given this distinction, we can bring back each Anti-Cartesian proposition to one of these cases.¹

In our opinion, there are at least three reasons for preferring it:

1. It is more grounded. Our distinction is based on an analysis that, although only sketched (and probably still incomplete), it is grounded on theoretical bases (specifically, on the rule played by each property of knowledge in generating Anti-Cartesianity), whereas Tennant does not justify his distinction in any way.

2. It is inclusive of Tennant’s distinction. We could settle Tennant’s examples inside our frame and give an explanation of their unknowability:

   - About inconsistencies. Even though they could be derived in two different ways (1.2 and 3.1b, 3.2b), we include them in kind I. Our distinction is able to show that (Fact) is sufficient for explaining their Anti-Cartesianity.

   - Propositions whose form is (p & ¬Kp) are included in kind II. Their form is a particular case of (p & q), where q ├ ¬Kp.

   - About “No thinkers exist”. The advantage of our distinction over Tennant’s is clear. We should not recall Hintikka’s notion of existential inconsistency for considering the above proposition unknowable. Existential inconsistency is not the reason for logical unknowability (Anti-Cartesianity) of the above proposition, but vice versa; the existentially inconsistent proposition is unknowable because, on condition that the existence of thinkers is necessary for knowing, it is a proposition p, such that p ├ ¬Kp (sub-case 1.1). A check of the validity of this explanation is the fact that if we imagine (per absurdum) that knowing is possible without existing, then existentially inconsistent propositions are dead knowable.

   Given our analysis, the references to concepts like Hintikka’s existential indefensibility in explaining Anti-Cartesianity seem to be misleading. Anti-Cartesianity grows only from knowledge and its properties and other propositions, such as those that are existential indefensible, are indirectly unknowable, only insofar as they are logically unknowable (i.e. reducible to

¹Note that Kind I might be reduced to Kind II, because the analysis of (Fact) could be included inside that of both the properties. In fact, in case 3, we have found all of the sub-cases of 1 repeated. But, the distinction of the two kinds is not endangered by this: in fact, our distinction is based on the sufficiency or insufficiency of (Fact) for having Anti-Cartesianity.
exploited cases). In this sense, our analysis is independent from non-epistemic notions.

3) *It is more complete.* Our distinction can include examples of propositions that, in Tennant’s distinction, would not find a place or would stay in the middle between two kinds: the above reported, problematic “button-sentence” is an example of the former case. The proposition “I pushed on the button”, given that if one pushes on it she forgets doing so, is a proposition with this property: $p \rightarrow \neg Kp$; and so, it is a proposition $p$, such that $p \vdash \neg Kp$ (sub-case 1.1). The proposition, not considered in Tennant’s distinction, is here in *kind 1*.

An example of a proposition in the middle of two of Tennant’s kinds, is “there are not known propositions”. In Tennant’s distinction, this proposition could be included in the third kind: “proposition that $\phi$ is known may be logically inconsistent because of its own overall logical structure, involving iterations of $K$”¹, or in the second kind: “knowledge of a (consistent) proposition $\phi$ may be impossible because the very act of considering or judging (falsely) that $\phi$, requires the falsity of (some consequence of) $\phi$”. In our distinction, the above proposition is Anti-Cartesian of the first kind because it is a proposition $p$, such that $p \vdash \neg Kp$, deducible with (Fact) alone.

V. Conclusion

Some final comments on the analysis given in section II: as shown in the last section, such analysis can be taken as the basis for a new distinction of Anti-Cartesian propositions, which is better motivated and more complete than Tennant’s distinction. But the analysis could also be interesting for other reasons, which are independent from the purpose of classifying such kinds of propositions. For example, if complete and correct, the analysis gives the way to form a complete list of propositions, problematic for the supporters of an epistemic notion of truth.

In our opinion, our analysis should be useful to an anti-realist arguing for the *Knowability Principle*, at least in terms of understanding more deeply the reasons for some of its problems. However, we do not want to consider our results as a criticism to the anti-realist’s position (at least, no more than the Knowability Paradox does). On the contrary, they could also be used for analyzing some features of the Anti-Cartesian propositions and for suggesting some other restrictive solutions to the paradox, which are better motivated and more complete than Tennant’s.²

---

¹By “there are not known propositions” is derived an inconsistency for the iteration of $K$ as follows:
1) $K(\forall p (\neg Kp))$ (Ass.)
2) $\forall p (\neg Kp)$ by (1) and (Fact.)
3) $\neg K(\forall p (\neg Kp))$ by (2), substituting (2) in (2), (1) and (3) are inconsistent between them.
²Rosenkranz (2008) gives an example of a more complete restriction, which is able to solve some further problems of Tennant. Our results could be useful for a further improvement of the restriction.

192
Logically Unknowable Propositions:  
A Criticism of Tennant’s Three-Partition of Anti-Cartesian Propositions

References


Available at http://paulegre.free.fr/Papers/fitch_rev.pdf.

Symbolic Logic 28: 135-142.

of Philosophy 77: 422-428.

Hintikka, J. (1962a). “Cogito, Ergo Sum: Inference or Performance?”. The  

Cornell University Press.


Hope”. In Salerno, J. (ed.), New Essays on the Knowability Paradox, Oxford:  
Oxford University Press.


59-75.


Imagine someone, Jones, who loses the ability to speak and understand English; later he recovers it. Does he know English during the time the ability is lost? According to Noam Chomsky, he does. Jones's recovery requires that he know English and have access to this knowledge. Now imagine someone exactly like Jones, call him Smith, who dies without recovering his English. Chomsky claims that Smith's knowledge also persists—admitting that we would want additional evidence to support this claim.

... other evidence (say, from autopsy ...) might lead us to conclude that Smith ... nevertheless retained his knowledge of English intact after having completely lost his ability to speak and understand (Chomsky 2000, 51).

In short, recovery of the lost ability rests upon possession of knowledge of language intact.

This strikes me as quite remarkable. How could someone who never speaks or gives any evidence of understanding what anyone says retain “his knowledge of English intact”? At the same time, the case isn't entirely far-fetched. If the only difference between Jones and Smith is that the latter dies before he recovers, it seems unreasonable to deny him the knowledge we grant to Jones, who has the good fortune to live. If we look at language as a part of our physiological make-up—something we might think of as the language system or language organ, as the reference to evidence from autopsy clearly implies—it seems reasonable to credit both Smith and Jones with the same knowledge. If Smith's physiological constitution remains intact during the coma from which he never recovers, then he's no different from Jones with respect to his physiology, and if, further, some system of organs is language, we might well say that he retains his knowledge of language even during the silent months or years.

But is this reasonable? Chomsky's analysis of knowledge of language really isn't analogous to the uncontroversially physiological. If the immune system completely stopped producing antibodies but later recovered, we wouldn't say that in the interim “the immune system is present intact”. That the physical
components which make up the immune system are still present isn't the same as the system's being intact. Knowing English doesn't mean that some organs are present—it means that you do or can do certain things.

In this paper, I argue that these concerns cannot be ignored; they reveal a fundamental problem with the theoretical basis of Chomskyan linguistics. In order to show this, I argue that the Chomskyan Revolution in linguistics—Cartesian Linguistics, as it is often called—carries within it the seeds of what Ryle called a Category Mistake.

Actually, there are two Category Mistakes. The first is close to the one Ryle uses against the “Official Doctrine”; it lies in Chomsky's switching from physiological talk to mental/knowledge talk. The second lies in his appeal to the methodology of the natural sciences as a justification for claims about the nature of our knowledge of language. Both qualify as 'theoretically interesting category-mistakes” because 'they allocate concepts to logical types to which they do not belong' (Ryle 1949, 17).

The Revolution

The roots of Chomsky's Cartesian Revolution lie in 17th century philosophy and linguistics. In addition to Descartes, Chomsky cites the writings of Wilhelm von Humboldt, Gottlieb Leibniz, John Locke, Friedrich von Schlegel and S. T. Coleridge. The outlines of this Revolution can be summarized in the following two claims.

1. Cartesian Mentalism. Language can only be explained in mentalistic terms, and it rests on propositional knowledge which we have innately, or as a species trait.
2. Galilean Naturalization. Linguistics is a natural science; as such it must follow the methodology of the natural sciences and be held accountable to the same standards.

By Chomsky's own reckoning, his revolutionary change to linguists lies in his rejection of the Empiricist tradition, along with its 20th century progeny, behaviorism. The Empiricist approach to understanding human language favors studying only what we can observe through the five senses without filtering these observations or favoring any of the data over others. In place of this, Chomsky advocates a return to Rationalism. The Rationalist approach requires that we go beyond what is observable and look for something underlying what we observe. In the case of cognition in general and language in particular, this entitles us to talk of the mental, as distinct from the physical, and to sort the available data of observation so that some of the data can be discounted and others given more weight in our calculations. To see the implications of this

---

1Another possibility is that we have innate knowledge of a theory of language, or competence; however, for this paper, it doesn’t matter what we know innately, just that we know something related to language innately.
The Chomskyan Revolution: A Category Mistake?

Revolution, I examine each of these claims in more depth.

*Cartesian Mentalism*

This aspect of the Revolution can be approached through the following question. Should linguistics be seen as contributing to the study of the mind or should it be seen as collecting a variety of facts about languages as they are used in various settings, by various peoples, and as part of social structures? Chomsky argues that treating language as just one of many of the “tools” invented by humans to help negotiate their way around the world has failed to produce any true understanding of language and its role in human cognition. If, by contrast, we view language as essentially mentalistic, the linguist can uncover the explanatory roots of language in the structure of the human mind, and at the same time, produce a greater understanding of the structure of human knowledge (cognition).

Up to the 1950s, linguistic research followed the traditions of the social sciences, psychology in particular, and eschewed talk of the mental or the mind. Since the “mind” is not accessible for empirical investigation, anything one says about it can be nothing more than idle speculation based on some inherited and primitive faith in man’s superior standing in nature. As a result, research into the nature of language consisted in producing structural descriptions, with a heavy emphasis on the ways in which languages differed from one another. The anthropologist often led the way in linguistic research because he looked to the social institutions developed by humans in pursuit of their place in the world, and language was preeminent among these social institutions.

In Chomsky's view, the methodology of this strict empirical process of data gathering and generalization created two problems.

1. It makes it impossible to arrive at an *explanation* of language based on the research; and
2. It rests on the rejection of Descartes's central insight into the nature of human language, an insight which seems firmly rooted in any unbiased examination of language and its functions.

... human language, being free from control of identifiable external stimuli or internal physiological states, can serve as a general instrument of thought and self-expression rather than merely as a communicative device of report, request, or command (Chomsky 2002a, 11-12).

In place of the Empiricist model, Chomsky advocates adopting an approach which not only allows for explanation, but which demands explanation as the aim of linguistic research. In the case of language, this explanation must be mentalistic.
A theory of linguistic structure that aims for explanatory adequacy incorporates an account of linguistic universals, and it attributes tacit knowledge of these universals to the child....The important question is: What are the initial assumptions concerning the nature of language that the child brings to language learning, and how detailed and specific is the innate schema (the general definition of “grammar”) that gradually becomes more explicit and differentiated as the child learns the language? ... Consequently, the task of linguistic theory must be to develop an account of linguistic universals that ... will not be falsified by the actual diversity of languages and,... will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars ... (Chomsky 1965, 27-28).

The empiricist model forces us to look for explanation in the data available to our senses: physical and behavioral manifestations of linguistic responses to features of the observable environment. In his review of B. F. Skinner's *Verbal Behavior* (Chomsky 1959), Chomsky gives a devastating critique of this approach, deftly illustrating its shortcomings and the petitio that lies at its heart.

By contrast, the explanation we seek cannot be found by looking at the surface, which reveals only differences. The linguist should be interested in what binds different languages together, and for this she must turn to the deep structure, the inner side of language, which is “...a simple reflection of the form of thought” (Chomsky 2002, 75). In other words, we look to the mental; here, and here alone, lies the explanation of language.

The central doctrine of Cartesian linguistics is that the general features of grammatical structure are common to all languages and reflect certain properties of the mind (Chomsky 2002a, 94).

In order to solve the problem of Linguistic Creativity, early linguists recognized the need for a theory of Human Understanding, which would yield a universal grammar, a theory of language's universal features. Believing that they could not produce such a theory, linguists turned to philosophy for the solution. Modern linguists, Chomsky contend, have simply ignored these issues.

Although it was well understood that linguistic processes are in some sense “creative,” the technical devices for expressing a system of recursive processes were simply not available until much more recently. In fact, a real understanding of how a language can ...“make infinite use of finite means” has developed only within the last thirty years, in the course of studies in the foundations of mathematics. Now that these insights are readily available it is
possible to return to the problems that were raised, but not solved, in
traditional linguistic theory, and to attempt an explicit formulation of
the “creative” processes of language. (Chomsky 1965, 8)

Now that we are able to represent recursive processes within a formal
system, we can account for the problem of Linguistic Creativity without
awaiting some philosophical account of the mind.

Chomsky's Cartesianism consists in making this formal theory part of the
stock of human knowledge, and, as such, available to human understanding.
The fact that human understanding is structured so that it includes this theory
of language provides the long sought explanation of language, along with the
solution to the problem of the structure and nature of Human Understanding.
Thus, the problem which early linguists had turned over to philosophy has been
resolved by linguistics itself.

Galilean Naturalization

Having shown that the proper goal of the linguist's study of language is the
discovery of what is universal to human languages, and, moreover, that only a
mental account can produce an explanation of language which accounts for the
fact that all of the apparently different and varied languages share a core of
features, Chomsky now turns his attention to the question of how we should
conduct our study the mind and language.

Like linguists before him, Chomsky advocates a scientific approach to the
study of language: the linguist's primary task is the collection of data through
empirical observation and experimentation. Unlike earlier linguists, however,
he takes the point of this data collection to be uncovering an explanation of
human language which rests on species-universal features of the mind. To
understand human language, we must look for something which all humans
have in common.

This leads him to examine the nature of linguistic data, and how we go
about collecting and analyzing them. The data are, Chomsky, argues not in
service of a theory about a social phenomenon, they are in service of a theory
about a natural object. Therefore, the data the linguist uses are the same sort of
data a biologist or chemist uses, not the sort of data used by anthropologists or
sociologists. In addition, the data are not collected to produce a description of a
set of natural objects, but to produce a theory to explain the unity of this set.
The linguist is now seen as a natural scientist, not a social scientist, and her
goal is the production of a theory—a theory of linguistic competence—which
will serve to guide future research programmes in just the way that the
physicist's theory guides his future research design.

The goal—whether one be a child learning language or a linguist engaged in
high-level syntactic theory—is to sort the data in such a way that one can
produce a formal theory which meets the criteria of consistency, elegance and
optimality. This is expressed in the Axiom of Categoricity, which “[makes]
our linguistics a kind of mathematics within which inconsistency is by
definition impossible” (Joos 1950).

This isn’t, however, all. In Aspects (Chomsky, 1965), Chomsky writes that the Generative Grammar is “psychologically real”; in On Nature and Language (Chomsky 2002b), this has changed a bit.

In natural language there is something in the head, which IS the computational system. The generative system is something real, as real as the liver; the utterances generated are like an epiphenomenon\(^1\) (Chomsky 2002b, 110).

However, he uses the expression “mind/brain” to signal that he remains committed to mentalism. Whether it’s in the “mind”, the “brain” or the “mind/brain”, the important point is that language is to be studied just like the liver. This is what it means to make linguistics part of natural science: Language is a natural object, and should be studied as such.

Leaving aside the question of whether language is really a natural object and not a social phenomenon, something which is not as clear cut as Chomsky would have us believe;\(^2\) his proposal to treat the study of language as a scientific enterprise, subject to the standards of our most rigorous scientific investigations, seems straightforward and reasonable. There are, however, some peculiarities embedded in the shift of linguistics away from the social sciences and into the natural sciences. This aspect of the Revolution, Galilean Naturalization, is the claim that language is a natural object, something which is a feature of the human species, in the way that gills are a feature of fish; but that is not the end of it.

Chomsky turns to the methodology of the natural sciences. He takes it that the natural sciences have a set of practices and standards associated to them, including the idea that data are gathered in service of testing and confirming or disconfirming a theory. One begins with a set of assumptions and tries to discover whether the data support these assumptions or require that they be modified or rejected entirely. Data are always examined from the perspective of a theory, and are used to construct a theory, which may be a modification—or at the extreme, a rejection of—the theory we begin with.\(^3\)

We begin with puzzles about unexplained phenomena, which we try to sort out into categories, caring little about boundaries, and not expecting the categories to survive inquiry (Chomsky 1996, 32).

\(^1\)His opposition to behaviorism is stronger than ever. Behaviorism not only fail to provide an explanation of language, the current criticism says that it studies something which isn't even part of language.

\(^2\)Ferdinand de Saussure certainly would argue against this, and modern sociolinguists offer a different perspective on the relationship between social institutions and language than that offered by Chomsky. See Hanna 2006 for a discussion of this.

\(^3\)For the sake of argument, I simply accept Chomsky's characterization of scientific methodology.
This method of seeking an explanation, not a description, no matter how detailed, through theory construction based on data which are filtered through the theory is the so-called “Galilean style”.

... the Galilean style referred to that major change in the way of looking at the world: you’re trying to understand how it works, not just describe a lot of phenomena, and that’s quite a shift (Chomsky 2002b, 100).

Galilean Naturalization has, therefore, two parts.

1. Classifying linguistics as a natural science—Naturalization.
2. Advocating a specific methodology for conducting research—the “Galilean style”.

Pace Chomsky, they do not entail one another. One might accept one and reject the other. I don't discuss this, but look at Chomsky's claim that to produce an explanation linguists must accept both.

The Galilean style ...is the recognition that it is the abstract systems that you are constructing that are really the truth: the array of phenomena [The data! ] is some distortion of the truth because of too many factors, all sorts of things (Chomsky 2002b, 99).

The goal of linguistic research is the production of a formal theory which provides an explanation of language. The research is conducted with the understanding that language is as much a natural object as the Grand Canyon or the liver, and as in all natural sciences, it is the theory of this object which is “really the truth.” In light of this, it is fair to say that for Chomsky, theory construction follows the model of mathematics and the theory is subject to the requirements of consistency and elegance which apply to all axiomatic systems.

**Revolutionizing Linguistics**

What's wrong with it? What's wrong with supposing that one conducts experimental research into natural objects in order to produce theories which are formal and abstract, and that these theories inform future in the form of data sifting, as well as being informed by the data when we modify the theory? Isn't this, more or less, what scientists (even social scientists) do? Adding mentalism may be little more than a façon de parler, nothing rising to the level of Category Mistake. Why do I suppose that the Revolution divides itself along logical lines which amounts to a category mistake? An article by Max Tegmark in *New Scientist*, September 15-21, 2007 contains the clue to my charge.

Tegmark, a professor of physics, is interested in answering this question:
What is the universe made of? His brief article isn't about language or the issues surrounding linguistic research, however, it contains some interesting lessons for these issues.

What is the Universe really made of?

Tegmark argues that “...our universe is not just described by mathematics—it is mathematics”, a view which he describes as an “extreme” version of a view tracing from Pythagoras through Galileo to modern physics’s enormously successful use of mathematics. If “there exists an external physical reality independent of us humans” (38), how better to account for the success of physics than by identifying the world with the theory itself? Since the theory is mathematical formulae, so too is reality. Like Tegmark, Chomsky is a realist, holding that the goal of linguistics is to describe a part of the “the one and only world”; and Chomsky takes physics as providing the standard for scientific research. Thus, it is not a distortion to look at this theory in Tegmark’s terms.

Theories of physics have traditionally contained two parts: “[1] mathematical equations, and [2] words that explain how the equations are connected to what we observe and intuitively understand” (numbering added, 38). [2] is where concepts such as proton, stars, and molecules are introduced, while these may be convenient, Tegmark argues that since we humans have invented them, they are of necessity imprecise and confusing. According to him, [2] is “baggage”, if we could approach the universe directly through mathematics, as a supercomputer would, we might find “a description of external reality that involves no baggage” (38). Tegmark advances a Platonic view of mathematics: the notational devices are invented by humans to mark out mathematical equations which present mathematical structures which “we discover ... and invent only the notation for describing them” (39). If you are a metaphysical Realist, then you must accept his view that “our physical reality is a mathematical structure”, we “live in a gigantic mathematical object” and “[e]verything the world is mathematical—including you” (39).

Normally, one might take this as an elaborate, if somewhat precious, way of saying that the universe and everything in it can be modeled or described by a system of mathematical formulae. However, given Tegmark’s stated disdain for ordinary—that is to say, the human languages we use for communicating with one another—language’s ability to describe the world, it is clear that he means more than this. Ordinary language masks reality, mathematics presents it in all its truth and glory.

---

1Chomsky’s commitment to some form of metaphysical Realism is found throughout his writings, this phrase appears in a number of places, including Chomsky 1996, 45.
2The idea that natural language disguises the “true” form of language is familiar to philosophers.
The Category Mistakes

The Chomskyan Revolution fits this picture. The formal system produced as the result of Galilean Naturalization and the Axiom of Categoricity constitute the formal system of [1]. Chomsky's description of Galilean Naturalization fits the way work in syntax is done, as well as fitting many of Chomsky's claims about the nature of syntax and how it is to be evaluated.¹ If Tegmark is correct and the natural sciences are aimed at producing formal, mathematical theories, Chomsky's Galilean Naturalization fills the bill: it promises a formal theory of language on a par with the formal, mathematical theories of physics.

Now Cartesian Mentalism emerges as a thesis about how to interpret that research—its theories, its experiments, and the very data it uses. Cartesian Mentalism is Chomsky's attempt to describe the nature of the theory produced under Galilean Naturalization in ordinary terms.

The interpretation under Cartesian Mentalism provides a compelling and appealing picture of language and its role in human development, as well as giving insight into how learning occurs and how the human mind functions. But is it the only interpretation? And is it the only interpretation that produces an explanation of language and the mind?

The answers to these questions are far from self-evident, but it's safe to say that they are most likely negative. Even if we assume—following Tegmark's lead—that language is a logical system, there's nothing in that system that forces us to interpret it in the terms of Cartesian Mentalism. If we go a bit farther with Tegmark, we might be inclined to think that applying any interpretation—mentalistic, behaviorist, social or physicalist—is the real mistake. But this goes too far.

Pace Tegmark, interpretation isn't “baggage”, something to avoid at all costs. It's what gives a theory its content, what makes a theory about something rather than a mere formal system. Chomsky's mistake lies in taking the success enjoyed under Galilean Naturalization and inferring that this supports Cartesian Mentalism, ignoring the obvious truth that the interpretative framework provided by Cartesian Mentalism is one interpretation, but it is only one. There are others—behaviorism and social construct theories, to name only two—which are equally compatible with the formal system.

If Chomsky is to convince us that Cartesian Mentalism is the one and only correct interpretation, he needs to do more than say that the formal system produced under Galilean Naturalization proves Cartesian Mentalism. Cartesian Mentalism can neither prove nor be proved by Galilean Naturalization, they aren't comparable in the same way that coming home in a flood of tears and coming home in a sedan chair aren't comparable. It is altogether likely that the theory of language produced by Galilean

¹It would also account for claims to the effect that the goal of syntax is to produce a theory of language and a theory of the brain, which one of my colleagues has often used to describe his work in syntax. On this account, there is nothing significantly different between language and the brain—even if we call it the “mind/brain”—and the theory we aim to produce is a formal, axiomatic system.
Naturalization is compatible with Cartesian Mentalism, but that's hardly enough to force us to accept Cartesian Mentalism and eschew all forms of behaviorism, and all interpretations of language which rest on the notion that language is a social phenomenon. Especially when one remembers what entitled Smith to his knowledge of language.

... other evidence (say, from autopsy, were enough known about the brain sciences) might lead us to conclude that Smith ... nevertheless retained his knowledge of English intact after having completely lost his ability to speak and understand (Chomsky 2000, 51)

Having produced a theory which accounts for language in terms of some organic system, Chomsky' follows his Cartesian Mentalism and treats this system as mental. He then concludes that knowledge—what he calls our “ordinary” conception of knowledge—survives so long as one's organs survive. The Category Mistake here is nothing more (or less) than the mistake which Ryle saw almost 60 years ago.

...The difference between the human behaviours which we describe as intelligent and those which we describe as unintelligent must be a difference in their causation; so, while some movements of human tongues and limbs are the effects of mechanical causes, others must be the effects of non-mechanical causes, i.e., some issue from movements of particles of matter, others from the workings of the mind (Ryle 1949, 19).

Chomsky may dress it in more modern language, but it comes to the same thing.

This is a mistake involving interpretations, the second part of a natural science, according to Tegmark; it is confined to [2]. There is another Category Mistake which involves both parts of a scientific theory, this bridges [1] and [2]. Cartesian Mentalism neither proves nor is proved by Galilean Naturalization because they aren't comparable in the way that coming home in a flood of tears and coming home in a sedan chair aren't comparable. Yes, one may come home in both, but this doesn't erase the logical differences between them. The theory produced by Galilean Naturalization is compatible with Cartesian Mentalism, but it's also compatible with behaviorism, and an understanding of language as a social phenomenon.

If we take seriously the notion that natural scientists construct formal theories to explain nature, and that these theories are then tested under various interpretations, the flaw at the heart of the Chomskyan Revolution lies in welding an ontological claim—Cartesian Mentalism—onto a methodological claim—Galilean Naturalization. In doing so, Chomsky imposes his own version of ontological Realism onto a model of empirically based research by

---

1See “Language and interpretation” (in Chomsky 2000, 46-74), for a full discussion of Chomsky's claims about the nature of knowledge of language.
The Chomskyan Revolution: A Category Mistake?

fiat, and draws the conclusion that the two are mutually supportive. In fact, they are entirely different enterprises. One—Galilean Naturalization—is fought out in the laboratory; the other—Cartesian Mentalism—will be decided in the halls of the academy by theoreticians, be they philosophers, linguists, physicists, or psychologists.

References

Kurt Gödel, as is well known, once offered an explicit analogy between how we acquire knowledge of mathematical objects (construed as platonic entities), and how we acquire knowledge of the external world (Gödel, 1983a, 483-4):

‘But, despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don’t see any reason why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception...The set-theoretical paradoxes are hardly more troublesome for mathematics than deceptions of the senses are for physics.’

Gödel (1983a, 484) is careful to note that this mathematical intuition is not to be understood as providing immediate (i.e., non-inferential) knowledge of mathematical objects:

‘Rather it seems that, as in the case of physical experience, we form our ideas also of those objects on the basis of something else which is immediately given. Only this something else here is not, or not primarily, the sensations.’

From the view that our knowledge of physical objects is a matter of some kind of inference from a ‘given’ (appearances) which is not itself knowledge of physical objects, Gödel concludes that the question of the existence of mathematical objects is ‘an exact replica of the question of the objective existence of the outer world’. Gödel’s argument, then, for the claim that there obtains a parallel between the two forms of skepticism (mathematical and external world skepticism) may be formalized as follows:

1. Knowledge of the external (physical) world is dependent on reliable sense experience.

---

1Emphases in original.
2. Sense experience gives us information about how the world appears.

3. In order for knowledge of the external world to be possible, it must be possible for us to infer such knowledge from what is given via sense experience (appearances).

4. Knowledge of the mathematical realm is dependent on intuition (which is something like sense perception).

5. Intuition gives us information about how the mathematical realm appears.

6. In order for us to have mathematical knowledge, it must be possible for us to infer how things are in the mathematical realm from what is given via mathematical intuition (appearances).

7. Skepticism about a type of entities arises whenever there is a gap between how those entities appear to us, and how they are in reality (where ‘how they are’ is understood so as to include the question of whether they actually exist). [implicit assumption]

8. This gap is present in both Knowledge of mathematical objects and knowledge of the external world.

9. Knowledge of mathematical objects and knowledge of the external world are equally vulnerable to skepticism.

Moreover, the way round skepticism, for Gödel (1983b, 456-7), is in both cases to make a pragmatic appeal to the needs of satisfactory theory construction:

‘...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. They are in the same sense necessary to obtain a satisfactory system of mathematics as physical bodies are necessary for a satisfactory theory of our sense perceptions and in both cases it is impossible to interpret the propositions one wants to assert about these entities as propositions about the “data,” i.e., in the latter case the actually occurring sense perceptions.’

Jody Azzouni (2000, 225-6), however, has argued that the analogy between skepticism about mathematical objects and external world skepticism is a weak one:

‘Meditation on how we sometimes take ourselves to be exercising faculties of observation when we are not (e.g., dreaming) leads us to wonder whether we can ever be sure of what we are (presumably) observing. The sceptic warns us that when we think that we are using our eyes to observe, we may not, in fact have eyes at all; that is, we may not be using the actual epistemic methods we think we’re using, but merely thinking we’re using them.’
If, however, in opposition to the sceptic, this epistemic story is left intact, and we ask what would happen if all the objects around us disappeared, the answer is obvious. We would notice something was missing ("where’d everything go?" we would wonder plaintively).¹

This, Azzouni urges, is in stark contrast to the mathematical case. Here, Azzouni claims, if mathematical objects ceased to exist, mathematical practice would continue “as usual” (Azzouni, 2000, 225). Moreover, Azzouni (2000, 226) argues that this thought:

‘... is not a skeptical one, nor can it be answered naively by simply pointing to day-to-day mathematical practice, and claiming that just as ordinary empirical practices in the sciences justify scientific pronouncements, so too, and to the same extent, ordinary mathematical practices justify mathematical pronouncements. For day-to-day mathematical practice offers no epistemic role for mathematical objects, and so it does not respond to the worry that there are no mathematical objects for its theorems to be true of. It is compatible with mathematical practice as it is (we really are using computers, using pens, carrying out proofs) that, despite this, there are no mathematical objects.”²

The claim then, is that the lack of an epistemic role for mathematical objects serves to weaken the analogy between external world skepticism and mathematical skepticism: we are not in exactly the same dire epistemic straits with respect to our knowledge of the external world as we are with respect to mathematical entities, and the platonist cannot deflect criticism of his epistemology by urging that platonists and believers in the external world are all in the same epistemic boat with respect to the entities they posit.

The terms of Azzouni’s thought experiment are not uncontroversial. J.R. Brown (1990 107), for example, holds that if mathematical objects disappeared, then we would no longer have intuitions about how mathematical reality appeared to us. If this is correct, then in the absence of mathematical objects, mathematical practice, contra Azzouni, would not go on ‘as usual.’ Bob Hale (1990, 128) holds that such thought experiments are incoherent, as they ask us to suppose the non-existence of objects that exist necessarily. These objections, however, do not seem fatal: there seems nothing incoherent about Azzouni’s supposition that mathematical objects might fail to exist (indeed it seems merely a colorful way of introducing the claim that such objects play no epistemic role in the production of our mathematical beliefs); and we might wonder how Brown knows that if mathematical objects disappeared, our intuitions would accompany them.

Azzouni is, I think, correct to note the disanalogy: if we are roughly correct in our views about how our beliefs about the external world are produced, then

¹Emphasis in original.
²Emphases in original.
if we hold those methods constant, and if everything disappeared, we would indeed notice. This does seem to contrast with the mathematical case, Brown notwithstanding. But exactly which part of the Gödelian argument does Azzouni’s thought experiment attack? We can be confident that premisses (4)-(6), (8) and the conclusion, (9), would, given Azzouni’s nominalism, all be rejected by him. Not all are affected by his thought experiment, however. The experiment is silent with respect to mathematical intuition, so it cannot touch premisses (4) and (5); (7) is merely a statement of a general condition that is required in order for skepticism to arise; (8) is a consequence of (6) (together with the uncontroversial (3), and (7); and (9) follows from (8). Azzouni’s thought experiment, then, if it is to trouble Gödel’s parallel between the two forms of skepticism, seems best directed at (6): the claim that in order for us to have mathematical knowledge, we must be able to infer how things stand in mathematical reality on the basis of how this appears to us.

But there is a serious problem with seeing Azzouni’s thought experiment as undermining (6). In the absence of abstracta, mathematical practice may perhaps go on ‘as usual’, but unless this going on ‘as usual’ is understood so as to include the production of mathematical knowledge, then going on ‘as usual’ is not an essential part of the Gödelian argument. It is quite compatible with premiss (6) that mathematicians behave as they currently do without gathering knowledge (if, for example, there is no mathematical knowledge, due to their being no mathematical objects for us to have knowledge of. Mathematicians in this case would have lots of more or less justified, but false, beliefs). Thus, if going on ‘as usual’ does not include the production of mathematical knowledge, then the thought experiment does not affect (6), and Gödel’s parallel between the two forms of skepticism stands.

On the other hand, if going on ‘as usual’ is understood so as to include the production of mathematical knowledge, then the platonist will point out that Azzouni’s thought experiment begs the question against the Gödelian argument. For the platonist holds that if there were no mathematical objects, there would be no mathematical knowledge, and so (assuming, with the platonist, that we do now have such knowledge) we would not be able to go on ‘as usual’ (though we would, to ourselves, appear to). So, on this reading, the thought experiment does not support Azzouni’s conclusion, and once again, we must conclude that Gödel’s parallel between the two forms of skepticism remains, in its essentials, unscathed. Either way, Azzouni’s thought experiment seems to leave the main thrust of the Gödelian argument unaffected (though it does, of course, refute Gödel’s overly strong claim that the mathematical epistemological problem is an exact replica of the external world epistemological problem).

One might object to the foregoing that, when Azzouni claims that mathematical practice offers no role for mathematical entities, he is working with a formalist, rather than a substantive conception of mathematics—a view
Mathematics, then, is concerned with reasoning about clearly specified things or ideas. There is no reason why mathematical symbols should stand only for numbers (as in arithmetic, algebra, trigonometry) or for points (as in geometry). They can stand for anything. Whatever they stand for, we develop them according to the properties that thing has.’

Consider also the following remarks on George Boole’s conception of mathematics (Boyer, 1991, 578-9):

‘Here for the first time the view is clearly expressed that the essential characteristic of mathematics is not so much its content as its form. If any topic is presented in such a way that it consists of symbols and precise rules of operation upon these symbols, subject only to the requirement of inner consistency, this topic is part of mathematics.’

If we interpret mathematics along such formalist lines, then, the objection continues, there will be no reason to regard mathematicians as coming to believe falsehoods after the disappearance of mathematical entities (either because mathematical formulae are purely syntactic—i.e., meaningless—and so are not the kind of things that can be true or false, or because consistency in a mathematical theory suffices for truth). There will thus be no reason to deny that mathematicians can “go on as usual,” and so, Azzouni’s thought experiment stands. This, in turn, refutes Gödel’s attempt to save platonism by means of his analogy between mathematical and external world skepticism.

However, we need to consider carefully how this objection is supposed to proceed. In particular does the objection assume the truth of formalism to begin with? That is to say, does it assume that mathematical axioms are not made true in virtue of referring to and correctly describing platonic entities, but instead describe jointly consistent conditions for some relational structure, so that those axioms will be true of any objects (tables, chairs and beer mugs, to use Hilbert’s examples) satisfying the axioms of that structure? On pain of begging the question, we cannot assume this view of mathematical truth when arguing against Gödel (indeed, starting with this view would make Azzouni’s argument redundant—it is supposed to show that platonic entities play no role in mathematics; it can hardly show this if it assumes it to begin with). We must thus allow Gödel the claim that mathematical entities exist, and are the truthmakers for mathematical statements. If, then, the formalist objection is to avoid begging the question, it must recast Azzouni’s thought experiment in the following form. So long as mathematical entities existed, mathematicians came

---

I owe this objection, and the references that follow, to an anonymous referee; my thanks to him or her.
to believe truths about these entities, and what made their true beliefs true (and their false beliefs false) was precisely whether those beliefs correctly described those mathematical entities. However, one day, those entities simply ceased to exist. Mathematicians, following this occurrence, did not thereby come to believe falsehoods, and so (the formalist thought experiment continues), were able to go on as usual, contrary to what I have argued above.

The objection, however, cannot be maintained. To see this, note that what saves mathematicians from believing falsehoods in the revised thought experiment is that the nature of mathematical truth has changed. Whereas prior to the disappearance of mathematical entities, truth was determined by those entities, after the disappearance of these entities, truth is determined by whether certain purely formal structures are consistently described (so that any objects whatever satisfying the axioms describing such structures may play the role of mathematical objects). And this change in mathematical truth must occur if the objection is to avoid either (a) begging the question by assuming the falsity of Gödel’s view to begin with, or (b) falling foul of my earlier conclusion that, since mathematicians will believe falsehoods after the disappearance of mathematical entities, they cannot go on as usual. But, of course, if prior to the disappearance of mathematical entities, mathematicians came to believe statements whose truth or falsity was determined by the nature of those entities, and if—after the disappearance of such entities—they come to believe statements whose truth or falsity does not depend on the nature of any such entities, once again, they do not go on as usual (though, once again, no behavioral difference will occur). The attempt to save Azzouni’s thought experiment by moving to a formalist conception of mathematics thus cannot succeed.

Further, to resume the main line of thought, even if we grant Azzouni’s point that if mathematical objects disappeared, this wouldn’t make any difference to our mathematical practice, this hardly serves to undermine the main thrust of the Gödelian argument. For consider, the skeptical hypothesis used in the Gödelian argument is that, for all we know, it is possible that there is, right now, no physical world, but that we are deluded into thinking that there is one. This possibility obtains since all of the evidence available in the event of the existence of the external world would be present in the skeptical world. The mathematician is similarly situated re mathematical objects. In both cases, we are faced with beliefs about how things appear, and must somehow move from these beliefs to having knowledge of the entities in question. But if we are to arrive at any such knowledge (as opposed to more or less justified false beliefs), then those entities must exist.

Seen from this angle, Azzouni’s thought experiment begs the question against the Gödelian argument: it assumes that our belief-forming processes concerning the external world are right now operating reliably (i.e., giving us knowledge of the external world). This assumption is required by Azzouni’s argument, since his thought experiment requires holding the processes constant while the world disappears. Given this assumption, then if everything disappeared, of course we would notice. But the parallel between external
world skepticism and mathematical skepticism that is relevant to the Gödelian argument actually blocks this assumption. This parallel has to do with the possibility that we may right now be deceived about what is causing those appearances on the basis of which we infer the existence of physical reality, on the one hand, and mathematical reality on the other. Given this parallel, we cannot, as Azzouni does, hold constant our epistemic story of how these appearances are caused, and then ask what would happen if everything disappeared. This move is at once ruled out by the parallel that operates in the Gödelian argument. We of course believe that we have knowledge of the external world, due to the reliable operation of our senses (similarly platonists believe that we have knowledge of the mathematical realm due to the reliable operation of intuition). However, the point is that the falsity of this belief is compatible with our having a phenomenologically indistinguishable experience. This possibility of our having phenomenologically indistinguishable experiences regardless of whether entities of a certain kind exist is, after all, the point of skepticism; and the point of the Gödelian argument is that this possibility obtains equally in the cases of mathematical and external world knowledge.

Bibliography


I take mental qualitative properties or qualia to be sensory or affective properties. As it is typically suggested, there are mental states that possess qualia. Mental states that possess those properties are perceptual experiences such as those involved in seeing red or tasting liquor, bodily sensations such as feeling an itch and felt moods or emotions such as feeling elation, grief, love or jealousy. These mental states have distinctive “phenomenological” characters, that is, felt or sensed qualities, by means of which they are usually identified as sensations of a certain type. It has become customary to refer to these sensed qualities of mental states in terms of which we characterise what it is like to have them as “qualia”\(^1\).

What is the ontological extent of qualia? As one might expect there is a wide range of views extending from David Chalmers’ (1996) panprotopsychism (roughly the view that an experiential element is present in everything that exists) to Daniel Dennett’s denial of the existence of qualia. Qualia have been paradigmatically seen as properties that are mental in the sense of being non-physical and non-physical in the sense of being essentially the objects of consciousness. In this sense, only non-physical properties can exemplify qualia. According to Chalmers (1996) for instance, a quale is an experiential/proto-conscious item and it is constituted out of the intrinsic natures of the fundamental physical entities. The intrinsic properties of the physical world include experiential or proto-experiential properties. Chalmers postulates experience as a fundamental property of physics alongside mass and charge and space-time. By contrast, according to Dennett (1991), ‘despite what seems obvious at first blush, there simply are no qualia at all’ (p. 520).

Now, relocation views are seen as middle ground attempts to naturalise qualia or to oppose the long-standing traditional association of their qualitative and the experiential character with the non-physical. With respect to qualia, relocationists attempt to move qualia out of the mind. According to one strain of this view, qualitative properties are properties of external physical objects. Colour qualitative properties for example, are moved all the way out onto the

---

\(^1\)E.g. *there is something it is like* for one to feel pain (i) as *there is something it is like* to fell an itch (ii) but (i) and (ii) refer to different types of sensations because the phenomenological/qualitative properties involved in those states are different: pain and itch respectively. Pains and itches are typical examples of qualia (and so is redness, greenness, hate, jealousy and so on).
surface of the perceived physical objects (Tye 1995, Lycan 1996, Dretske, 1995). On this view, when one perceives a red rose for instance, one is visually representing the actual redness of the rose. The represented redness of the rose is the actual redness of the rose itself. Thus redness is not a property of one’s experience but an externally constituted property of the perceived physical object. In this sense, qualia are out there, in the external world. The other strain moves qualia into the brain. This strain also has it that qualitative properties are not properties of the mind. A yellow or a red patch in one’s visual field for instance, is just a neurophysiological state in one’s brain (Block, 1999). I have argued elsewhere that the relocationist attempt to identify qualia with properties of external objects fails (Platchias, in press). I’ve argued that the standard arguments for such view, if they are successful, establish no more than a symmetrical supervenience relation between represented external content and qualia, and a supervenience relation alone (albeit symmetrical) doesn’t suffice for identity. In what follows, I will first outline why the representationalist cannot locate qualia ontologically and then suggest a way that could get us out of the predicament to locate qualia ontologically.

**Relations of Dependency and Identity**

Generally, physicalists find the supervenience relation useful. The key idea of supervenience is, crudely put, that there is no mental difference without a physical difference. One of the advantages of this relation of dependency between the mental and the physical is its clear implication that the mental domain is anchored in the physical domain without at the same time implying ontological reduction. To say that A supervenes on B is to say that if the properties of B (subvenient ones) are the same the properties of A (supervenient ones) cannot differ. Yet the supervenient properties need not be reduced to the subvenient ones (in the sense of them being *nothing but* the subvenient ones). In other words, the supervenience relation can safeguard the ontological autonomy of A relative to B. Concluding this section, we must also say that supervenience relation can also be symmetrical. Yet, symmetrical supervenience does not imply identity. From no ‘A-difference without a B-difference and no B-difference without an A-difference’ it doesn’t follow that A is the same as B. This is the main reason that representationalist views like

---

1I must note that the strain that moves qualia into the brain does not fair better. One of the reasons that most philosophers now think that the answer to the problem of qualia does not lie somewhere in the neurosciences is precisely the great difficulty of locating qualia ontologically. Neurophysiologically speaking, one of the main problems is that it appears that there is nothing intrinsic in the brain that constitutes the difference between a red and a green quale. It appears so to speak that there is no neurophysiological difference between these two states; to the best of our neurophysiological knowledge, that is, there are no anatomical differences in cells in the visual cortices that correlate with colour differences. The neurophysiologist V.G. Hardcastle (1995) has said for instance that the best neurophysiological theory might not be able to rule out a case of inverted spectra, namely a case where molecular/functional duplicates might nonetheless have inverted colour qualia.
Tye’s and Dretske’s fail to show that mental qualitative properties are one and the same as external properties of the environment. Representationalist arguments show at best that there holds a symmetrical supervenience relation between mental qualitative properties and external properties of the environment and this falls short from establishing identity. My monthly bank statement for instance, symmetrically supervenes on my payings in and takings out but it is not identical to those transactions.

It might be objected that the symmetrical supervenience of my finances and my bank statement is only contingent. What is needed for identity, the objection continues, is that the symmetrical supervenience holds as a matter of necessity; holds that is, in all possible worlds. But to see that even logically necessary symmetrical supervenience is not sufficient for identity consider the following. There is necessary symmetrical supervenience between propositions of kind K (true) and the facts they represent. There is for instance, symmetrical supervenience (logical) between the (true) proposition that the cat is on the mat and the fact that the cat is on the mat. But clearly the proposition (an abstract entity) is not the same as the cat being on the mat.

**Representationalism and the Location of Qualia**

In the discussion that follows, ‘representational content’ will be used as a convenient term of standing for what is represented. A representational property is a property instantiated by a particular as having it in virtue of its having representational content. A representational property of a representational state may thus be described as a pair composed of an attitude and a content. Generally, the notion of representation entails one thing’s standing for something else or itself (self-representation). Representing is not instantiating. Universals or objective properties (e.g. colours) are instantiated by particulars (or in the environment) as bodily sensations are instantiated in individuals. In general, the mind does not instantiate the properties it represents. An object’s shape, for example, is instantiated in the environment and it is represented in one’s perceptual experience.

According to Strong Representationalists, the qualitative character of our sensory experiences, that is, the apparent objects and properties of those experiences are merely representational, namely they comprise or contain the content of those experiences without thereby being actually instantiated in the mind.

Let us now draw a distinction between weak and strong representationalism:

1. Qualitative character supervenes on a certain kind of representational content (weak representationalism).
2. Qualitative character is one and the same as a certain kind of representational content. (strong representationalism)
An Anthology of Philosophical Studies

So when one perceives a red rose one is visually representing the actual redness of the rose. The represented redness of the rose is the actual redness of the rose itself. But what if the subject is hallucinating? Suppose that one is hallucinating a red rose and that there is actually no red rose in his visual field. In this case redness is an intentional inexistent: it is a property of the non-actual material thing (rose). The object does not exist but it is still an intentional object. If these considerations are on the right track then we learn something about the nature of qualia. And indeed the strong representationalist claims to have solved the big problem of locating qualia ontologically.

However, the nub of the representationalist/non-representationalist debate is the following two (supervenience) theses:

\(\alpha\) No difference in qualitative character without any difference in representational content.

\(\beta\) No difference in representational content without any difference in qualitative character.

In other words, two experiences with the same representational content necessarily have the same qualitative character and vice versa. If both \(\alpha\) and \(\beta\) hold then this is case of a symmetrical supervenience relation. Many philosophers have tried their best to think cases of experiences which are identical in their representational contents but which differ qualitatively (Block 1996, Peacocke 1983, Searle 1983). According to strong representationalists like Tye (forthcoming) no one as yet has managed to show that the proposition ‘where there is a qualitative difference there is a representational difference too’ is false. Let’s have a quick look at Peacocke’s case (1983). Suppose images of two trees at different distances represent them as having the same height while taking up differently-sized portions of the visual field. This case is thought to have suggested that there can be experiences alike in their representational content which differ qualitatively. But the representationalist reply is that these images actually differ in certain representational properties such as their representation of distance or of visual angle. Thus thesis \(\alpha\) is unrefuted (or so I will assume).

---

1A century ago, Franz Brentano argued that intentionality is the distinctive characteristic of mental phenomena and that every mental state is intentional. Ordinary relations cannot hold between existent and non-existent objects but the mind can contemplate non-existent objects and states of affairs. The non-existent objects needn’t be immaterial substances (Russelian sense data for instance), they can be non actual or non-existent physical things. This of course, is by no means uncontroversial. Prima facie, positing a non-existent intentional object to solve the problem of locating qualia ontologically will certainly raise eyebrows. When one has a hallucination of redness for instance, there is a certain actually existing property that one is aware of, in virtue of which his experience has the qualitative character that it has. I will not however, develop this point more and nor will I explore the range of the possible representationalist maneuvers here. My aim is to show that the strong representationalist has not managed to locate qualia ontologically on different grounds.
Alternatively, let us appeal to inverted spectrum scenarios i.e. cases where molecular/functional duplicates might nonetheless have inverted colour sensory qualities. One of the main advocates of such scenarios against representationalist accounts of qualitative character is Ned Block. According to Block (1998), the case of the inverted earth for instance, is such a scenario which amounts to the rejection of \( \beta \). I’m not going to discuss this argument here. For my present purposes the following suffices. Block here attempts to reject \( \beta \). But suppose that the representationalist has a plausible reply to the effect that Block’s argument cannot refute \( \beta \). It might be urged for instance, that Block’s example simply begs the question: no such inversion is possible and hence, the description of the example is incoherent. Thus let’s accept that \( \beta \) is unrefuted. Suppose further that \( \alpha \) and \( \beta \) are both true and therefore that there holds a symmetrical supervenience relation between the qualitative character of our experiences and representational content. Does this mean that they are identical? No. Symmetrical supervenience doesn’t entail identity. There are (other) strong representationalist arguments to be sure, but they all follow the same pattern. The typical reply is this: There is a phenomenal [qualitative] difference between the two cases, the representationalist will say, precisely because there is representational difference’ (Tye, forthcoming). However, I couldn’t possibly state here all possible representationalist arguments and their variations to prove this point.

The strong representationalist move here is to appeal to the thesis of the transparency of experience, namely to the idea that our experiences are transparent or diaphanous. Let us briefly see, what the transparency thesis is. According to strong representationalists (Tye 1995, 1998, Harman 1990) we normally see through perceptual states to external objects and we do not actually notice that we are in perceptual states. Visual experiences for example, are transparent to their subjects. Tye writes

Suppose you are facing a white wall, on which you see a bright red disk. Suppose you are attending closely to the colour and shape of the disk as well as the background. Now turn your attention from what you see out there in the world before you to your visual experience. Focus upon your awareness of the disk as opposed to the disk of which you are aware. Do you find yourself acquainted with new qualities, qualities that are intrinsic to your visual experience in the way that redness and roundness are intrinsic to the disk? Surely the answer to this question is ‘No’ (Tye, 1998, p.660, emphasis in the original).

The point is that when one turns one’s attention inwards to one’s experience of the certain features of the world, one is aware of the very same features. According to Tye, no new features of one’s experience over and above its

---

1See Tye (forthcoming) for an overview and discussion of the main arguments for strong representationalism.
being red, round, etc are revealed. We can of course, be aware of the fact that our mental state is representing them but our experiences are like transparent sheets of glass and therefore we are not introspectively aware of our visual experiences any more than we are perceptually aware of transparent sheets of glass. Now, since introspection reveals no new properties at all, all experiential properties (including qualitative properties) are external. As it turns out the main idea of the transparency thesis is that we have direct access (we are directly aware) of properties of external objects. We are not directly aware of any experiential properties. We are directly aware of intentional features of our experience, namely the objects and properties our experience is about and these are properties of external objects, not properties of our experiences. Introspection reveals no new properties (i.e. intrinsic to our experiences) in addition to those possessed by external objects.

Thus the transparency thesis is this: we are directly aware of only the external objects and properties that are represented in perceptual experience. Can this establish the truth of Strong Representationalism? Namely, the idea that the qualitative properties of our perceptual experiences are the same as the external objects and properties that are represented in perception? No. The transparency thesis clearly begs the question. What exactly is the evidence of introspection to tell us that introspectible properties (qualitative properties) are nothing over and above the (externally constituted) properties of external objects? Does introspection tells us that an object’s looking bluish to us is one and the same thing as the surface property of blue that the object instantiates? No. The claim that our introspection tells us that perceptual experience appears transparent or diaphanous means that objects have their colours and properties in general as we experience them. But from this it doesn’t follow that the external world has exactly those properties. What it does follow is that external world as we experience it has those properties. There is nothing in perception or introspection per se to show that what we are directly aware of are properties of the external world. Thus this argument does not show that objects have those properties independently of how we perceive them. I cannot rule out (via introspection) the view that an object’s looking bluish to me is a quality of my experience that I have in addition to my awareness of blueness as being the surface property of blue that the object instantiates i.e. I might be aware of the blueness of external objects in virtue of being aware of that property of my experience.

The only way that the strong representationalist can sustain the purported claim of the transparency thesis is to add to the evidence from introspection the assumption that there holds an identity relation between qualitative character and external properties or content. And maybe this assumption is a mistaken claim that depends on the conjunction of α) and β). Be that as it may, one

---

1There other problems with the transparency thesis, such as those presented by cases of illusion/hallucination and the fact that it seems that it is not at all clear that qualitative content does not outstrip representational content or that introspection does not reveal any qualitative properties which are intrinsic to our experiences (see Crane 2006 and Block 2003). I will however, assume that the representationalist has a plausible reply to those objections.
doesn’t use the transparency thesis to establish the identity relation, one uses identity to establish the transparency thesis. That is, the transparency thesis presupposes the very identity that purports to establish. Hence the transparency thesis is not independent evidence for identity and one cannot use the transparency thesis to establish identity.

Now, granted that our distinct qualitative properties resemble and differ from one another in ways that parallel the similarities and differences among perceptible properties of the corresponding physical objects of the environment and that our qualitative properties bear certain systematic relations to families of properties of physical objects and processes to which the organism can respond, we might wish to establish something like this: No change in how we experience the world with no change in the world and no change in the world with no change in how we experience the world. But this again is a supervenience thesis which however strong and albeit symmetrical fails to show that there holds an identity relation between external properties of the environment and qualitative properties. My conclusion then is that the most that the representationalist arguments can establish is the truth of $\alpha$ and $\beta$). However, what follows from $[\alpha$ and $\beta)]$ is that the qualitative properties of our mental states supervene symmetrically upon external properties of the environment and this fails to establish the identity relation that the strong representationalist must establish in order to locate qualia ontologically.

A Possible Objection

Since the symmetrical supervenience relation may not imply identity but it is implied by identity, it might be urged that it is still worthwhile arguing to this effect, namely to establish the supervenience relation because it is a necessary condition for the identity relation that the strong representationalist wishes to establish. Block for instance, is right to argue against the supervenience relation because if he’s right then the strong representationalist (the identity theorist) is wrong. But if he’s wrong then the identity theorist establishes a necessary condition for identity. The identity theorist however, still needs something more to establish identity. Maybe at this point, strong representationalists agree that the success of the strong representationalist arguments establish only a symmetrical supervenience relation but they go on to claim that we can still establish identity by way of inference to the best explanation. They might argue for instance, that those who object to the identity claim need to provide a reason for rejecting it, and showing lack of supervenience is usually the way.

But think of the following formula:

$$\Box \forall x \forall y (\Phi x \equiv \Psi y) \rightarrow \Phi = \Psi \ (1)$$

This formula reads: if necessarily, for all x and for all y, the property $\Phi$ of x is equivalent to the property $\Psi$ of y, then that property of x is the same as that
property of y. Thus if (1) is true it follows that symmetrical supervenience does entail identity. However, we cannot determine whether Φ is the same as Ψ from the mere fact that they are logically equivalent. Take the case of an equilateral triangle. In such a triangle, all three sides are the same length. Further, necessarily a triangle is equilateral if and only if it is equiangular i.e. all three angles are equal. Now, let Φ = equilateral and Ψ = equiangular. We can construct the following formula:

\[ \forall x \forall y (\Phi x \equiv \Psi y) \text{ or } \forall x \forall y (\Phi x \leftrightarrow \Psi y) \quad (2) \]

But clearly from (2) doesn’t follow that Φ = Ψ: equilateral is not the same as equiangular. But this is case where one could equally appeal to identity as an explanation on their being supervenient on each other, which otherwise appears to have no explanation. But that’s clearly false. And note that, as opposed to the strong representationalist case, this is a case of necessary symmetrical supervenience. Thus, one can turn the objection on its head and say that if strong representationalists want to establish identity then they must provide something more than merely offering (1).

But what could be the argument here? The only available argument to this effect seems to be the thesis of transparency of experience. But this claim clearly begs the question and can do no more than establish a symmetrical supervenience thesis. As I’ve argued, the only way that the strong representationalist can sustain the purported claim of the transparency thesis is to add to the evidence from introspection the assumption that there holds an identity relation between qualitative character and properties of external objects. But in this case, we don’t use the transparency thesis to establish identity we are using identity to establish the transparency thesis. Strong representationalists still lack a positive argument for identifying qualitative properties with any specific external property and in lack of such an argument I see no reason to think that identity is the best explanation. The non-representationalist or the weak representationalist has equal right to suggest that it is not. However, appealing again to the thesis of transparency of experience, we saw that introspection tells us that objects have their colours and properties in general as we experience them, not that the external world has exactly those properties. The external world as we experience it has those properties. Hence we are not and we may never be in a position to know whether our case is similar to the equilateral/equiangular triangle case.

So where are Qualia Located?

How are we to naturalise/ontologically locate qualia? Well, even if we fail to relocate them i.e. identify qualia with externally constituted properties of

---

1Notably, Occam’s Razor, namely the idea that we should not multiply entities beyond necessity is not always applicable. And in general, philosophy is characterised essentially by a quest for truth not by a n urge to get the simplest explanation.
objects (Tye 2000, Lycan 1996, 2001, 2006) or with neurophysiological states in the brain (Block 1999)\(^1\), we don’t need to eliminate them in order to naturalise them. Here’s how we might go about naturalising them. The main reason that qualia appear unamenable to a physicalist explanation is that they are mistakenly construed as intrinsic (in the sense of non-relational), atomic and unanalysable properties. If you couple ‘atomic’ or ‘simple’ property with the claim that it is irreducibly subjective (as opposed to objective or to what is amenable to a third person – scientific explanation) then there is nothing to get hold of when trying to explain such a property. This construal rests heavily on the mistaken assumption (which has acquired the status of an axiom) that the qualitative properties of our perceptual experiences such as a pain or an itch in one’s finger are essentially conscious. That is, if one is in a mental state with one of these properties then one is in a conscious state. According to most philosophers, although many types of mental states such as thoughts, desires and beliefs can occur unconsciously, our qualitative states/properties are essentially conscious. That is, if one is in a mental state with one of these properties then one is in a conscious state.

However, it appears that we have good reason to think that qualia can occur unconsciously and therefore reject the idea that qualia are essentially the objects of consciousness (and in that sense non-physical). Recent neuroscientific findings suggest dissociations between aware and unaware performance (Hollingworth & Henderson 2002, Hollingworth, 2003). Masked-priming experiments for instance, show that subjects see stimuli they are unaware of seeing. Though the subjects deny seeing the masked prime, it affects their subsequent behaviour and reasoning. Studies on blindsight (Weiskrantz 1986), visual-form agnosia and optic ataxia (Milner & Goodale 1995), change detection (Fernandez-Duque & Thornton 2000; Simons et al. 2002) and others provide similar examples. Fernandez-Duque & Thornton’s (2000) work on change detection for instance, has shown that changes in the environment can still affect performance even when they are not explicitly detected. Another example of a representation of a change of which the subject is unaware comes from Bruno Repp (2001). In his perceptual-motor synchronization task, participants tapped a key in synchrony with a sequence of auditory tones. In some trials, a change in tempo was introduced at a certain point of the sequence. Participants continued tapping after the end of the sequence, following which they reported whether they had noticed a change. For sequences in which participants reported being aware of the change, they

---

\(^1\) I said in the introductory section that according to this relocationalist view, qualia are physical properties of neurophysiological states in one’s brain, e.g. a yellow or a red patch in one’s visual field for instance, is just a physical property of a neurophysiological state in one’s brain. This view apart from striking one as unintuitive faces serious difficulties such as, for instance, multiple realisation of qualia and that there is nothing intrinsic in the brain that constitutes the difference between say, a red and a green quale; there are no anatomical differences in cells in the visual cortices that correlate with colour differences (see Hardcastle, 1995).
made successful adjustments (phase corrections). Phase corrections occurred even in trials in which the participants failed to report the change. This is taken to suggest that awareness of change is not needed for such correction.

Blindsight is referred to as an ability to respond appropriately to visual inputs while lacking the feeling of having seen them (while believing oneself not to be seeing them); in other words, when there is nothing it’s like for one to undergo these states. The work of one of the pioneer investigators of the phenomenon (Weiskrantz, 1986)\(^1\) and others suggests that the controversial and bizarre phenomenon of blindsight seems to establish that one can perceive things visually even in the complete absence of visual awareness. Blindsight patients may lack visual consciousness or awareness yet in forced choice paradigms they answer accurately questions about the shape, position and even colour of surrounding objects. Such people are unaware of the visual information they possess (even when they use it); they commonly take themselves to be guessing. One can plausibly then say that this dissociation between unconscious and conscious processing in blindsight cases suggests that there are perceptual states, which are not available to consciousness. It would follow that a blindsight patient lacks a certain kind of access to the content of those states.\(^2\) Blindsight patients even though it is believed that they store visual information (qualitative) they seem to lack a certain kind of direct access to the visual modality or to the corresponding perceptual content. And there are also cases such as when patients that say that they do not see anything on to their right side but if asked to guess what is there they guess with considerable accuracy. These patients have some sort of lesion in area V1 of the visual cortex but possess unimpaired function in the other distinct areas that the retinal image projects to.

It appears then that perceptual states can occur unconsciously. Discrimination tests suggest that both unconscious and conscious perceptual states can represent sensory features like colour. It appears that whatever distinguishes say red sensory states from blue sensory states is present even when the state is not conscious i.e. even if there’s nothing it’s like for one to be in that state. It seems then natural to describe these various subjects in mentalistic terms, as *seeing* the various properties of the objects, when explaining their accurate guessing or other discriminative behaviour, but without their being aware that they are so seeing. And the experimenters do so describe them, such descriptions feeding naturally into *psychological* explanations of the subjects behaviour. I propose we take such ascriptions of unconscious mental states as literally and realistically as we take ascriptions of conscious mental states in psychological explanations of ordinary behaviour\(^3\).

\(^1\)For a recent discussion see P. Stoerig, (1996).
\(^2\)This is something that almost all commentators on the subject agree on.
\(^3\)Think of cases of peripheral vision. Imagine for instance, that you are driving too fast to identify some obstacle lying on your path, but you adeptly manage to avoid hitting it. It is fair to say that unless you saw it you would have hit it. Since you didn’t hit it you saw it. But you failed to identify or notice it. Why think that in this case you didn’t literally *see* the object? And of course, it is widely accepted that emotions, feelings and desires can occur unconsciously.
The suggestion then is that qualia are not essentially conscious or essentially the object of consciousness. Consciousness is a matter of there being something it is like for the subject to be in a mental state. The mental qualitative states that we are not aware of (perceptual or sensory states) are not conscious: if one is not in any way aware of a mental state or that one is in that state then there is nothing it is like for one to be in it. And indeed, plausibly enough, if one is not aware of what it is like then there is nothing it is like for one. And this leads us to divide philosophical labour and to characterise qualia and consciousness (i.e. what it is like to have them) independently of each other and employ different explanatory accounts for them. This division of labour is the key to many difficult problems in contemporary philosophy of mind including spectrum inversion scenarios and the notorious explanatory gap. For instance, cases where molecular/functional duplicates might nonetheless have inverted colour sensory qualities –spectrum inversion scenarios- lose their intuitive force. If qualia can occur unconsciously then they can be defined in objective terms or in terms of functional roles. And if qualia are to be identified with functional roles then spectrum inversion scenarios don’t seem to be possible since functional roles cannot be inverted. Thus, to conclude, if qualia can occur unconsciously then they are amenable to a third-person examination; they will be amenable to scientific investigation. We saw for instance, that it is open to identify qualia with functional roles. This way we solve two problems at once: first we diminish the intuitive force of spectrum inversion scenarios and second we naturalise qualia avoiding the intolerable dilemma of relocation or elimination.

References


CHAPTER TWENTY TWO

Abstraction and the Interpretation of Quality

Donald V. Poochigian

Infinite divisibility denies reality of substance. Infinite regress denies reality of form. Both content and form endlessly approaching nothing, neither is made of anything, so neither exists. Yet being formal, infinite regress is self-contradictory. A formal proof of non-existence of form is self-defeating. Also being formal, infinite divisibility is self-contradictory. Form constituting relationship of content, if content is unreal because infinitely reducing to nothing, form is unreal.

Neither argument being possible unless form and content exist, when form and content devolve to nothing, both must be real, neither a construct. Presupposing form and content, both arguments show neither form nor content derive from some ultimate element or elements. Their existence can be only spontaneous, and thus inexplicable, both form and content being intrinsically immediate and primitive.

Thus a universal logic is self-contradictory. Everything presupposing logic (a rule of identity), logic exists independently of everything else. Anything reduces to other related things, and these into still other related things, and so on endlessly, rendering form real and substance unreal as Plato posited. But being a thing itself, form is unreal. Every rule of logic is divisible into related constituents identifiable by another rule of logic similarly divisible, etc. A fatal regress ensues in which there is no ultimate logic by which every other logic is identifiable. No logic being identifiable, there is no logic. Logic being relationship, there is no relationship. And things being constituents in relationship, there is nothing.

Yet things do exist, structure in reality being an emergent mystery. It is self-evidently immediate in experience, presupposing no prior standard. Certainly such a standard can be provided, but it is not incorporated in immediate experience. Coherent experience could not occur if it were incorporated because the standard of logic would be limitless. Logic is necessarily a self-evident intuitive reality, error only occurring within the context of another world, where self-evidence differs.

Truth occurs within such a reality, this sort of world constituting what exists. Incorporated must be a priori existents which are self-evident within a world, requiring clarification only when occurring within another world. There can be no truth in a world unless some truth is self-evident in it, and what is self-evident in one world need not be so in another world. Self-evident truth can exhaust a world’s content, or a world can contain a necessary truth.
An Anthology of Philosophical Studies

identifying conditional truth. Only assuming combined worlds need otherwise self-evident truth be derived from other self-evident truth to be veridical. A self-evident truth becomes a conditional truth within this different world.

No two objects are identical at least in identity and, qualitatively, in temporal and/or spatial location. Considering what makes it unique, any object is an analogical archetype—“a ‘base case’” in set theory—for a universal class of which it is sole member.\(^1\) As so, it is both intensional criterion and intentional constituent of the universal class of itself. It is intensional criterion because any member of the class is an analogical archetype for membership in the class. It is intentional constituent because any analogical archetype for membership in the class is a member of the class.

Identifying every element between a set’s limits, any set’s power set is ambiguous. Distinguishing each member of a set is understandable in contradictory ways, as generating both one indivisible thing and infinite indivisible things. Either every element is fused into one with nothing separating any from another, or every element is diffused into infinity with nothing linking any to another. Resolving the ambiguity is mutual identity of whole and parts.

How is it known fused whole is diffused parts and diffused parts are fused whole, since neither appears like the other? How is it known \(A\) if and only if \(B\) plus \(C\)? Identity of whole and parts is knowable only by self-evidence, each identifying the other in the dialectical circle of a corecursive hyperset.\(^2\) A complete set is proven by its containing all constituents of the complete set; and all constituents of the complete set are proven by containment in the complete set. Diffused parts and fused whole are intensional and intentional in turn, each criterion as limit for and instance as limit of the other.


\(^2\)Theorem 17.5 (Corecursion Theorem) Let GAMMA be a smooth operator, and let pi : C implies GAMMA ( C ). There is a unique map \(\phi : C\) implies GAMMA such that \(\phi\) satisfies GAMMA-corecursion for pi relative to any GAMMA-notation scheme for \(C\).

In other words, this tells us that there is a unique function \(\phi\) that we can compute in the following way:

1. Pick a GAMMA-notation scheme \(den : X\) implies \(C\).
2. For each \(c\) existing in \(C\), find its associated name ceiling brackets \(c\).
3. Apply over-bar pi to get over-bar pi(ceiling brackets \(c\)). In genera, this will contain the names of some elements \(c\) existing in \(C\), possibly including \(c\).
4. For each of these names ceiling brackets \(c\), find \(\phi(c)\) and substitute this value for ceiling brackets \(c\) in over-bar pi (ceiling brackets \(c\)).

We call Theorem 17.5 a Corecursion Theorem because \(\phi\) itself is involved in part (4) of this procedure. [Jon Barwise and Lawrence Moss, Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena (Stanford, California: Center for the Study of Language and Information, 1996) 250.] Alternately inducing whole from parts and deducing parts from whole, judging “a part cannot be a cause of a whole of which it is a part, nor can a whole be a cause of one of its own parts. It seems right to insist, in fact, that a cause and any one of its effects must be wholly distinct events, having no part in common,” is mistaken. [E. J. Lowe, A Survey of Metaphysics (New York: Oxford University Press, 2002) 163.]
Abstraction and the Interpretation of Quality

Indistinguishable and distinguishable whole being limits of any set, mutual identity as set limits occurs by iterative determination of fused whole and diffused parts. Each is identified by either consecutively resolving each disjunctive of the limits exclusively, or consecutively resolving each disjunctive of the limits inclusively. Exclusive disjunction of the limits generates diffused parts from fused whole; inclusive disjunction of the limits generates fused whole from diffused parts.

Within the same limits, any exclusive disjunctive iteration terminates in the same diffused parts, and any inclusive disjunctive iteration terminates in the same fused parts, after all disjunctives of the limits are resolved, identifying all set constituents. Constitutive of an axiomatic system, parallel computer architecture and macroeconomics manifest this structure. Mistaken is in circularity, “the information contained [has] no origin,”¹ because, “the sign for a function already contains the prototype of its argument, and it cannot contain itself.”²

Abstraction

Grouped and ungrouped qualitative elements being sensately indistinguishable, distinguishing them is abstraction. Experience composes two elements, quality and abstraction. Quality is content of sensation, and can be content of imagination.³ Abstraction is content of imagination. Quality occurs in time and/or space; abstraction occurs in neither time nor space. As content of sensation at least, quality is unimaginable in neither time nor space. As content of imagination at most, abstraction is unimaginable in either time or space. God exists in neither time nor space. So abstraction exists independently of time and space.

An abstraction is simple identity, being this and not anything else. Occurring in neither time nor space, it is indivisible. Logically complete thus, it is rational. Sensately undetectable, as shall be seen, a limited sequence is “sense,” and an unlimited sequence is “nonsense.” Abstractions being simple identities, 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.

Emergence occurs if all preceding disjunctives are exclusive, and submergence occurs if all succeeding disjunctives are exclusive. Supervenience occurs if some preceding disjunctive is inclusive, and

¹Lowe 340.
supervenience occurs if some succeeding disjunctive is inclusive. Emergent, submergent, and unchanging, abstraction is substantively inexplicable, illogical as such.\(^1\) Abstraction is formally explicable, logical as such, when self-evidently identified constituent of another abstraction or constituted by other abstractions.

An analogical archetype can act as a sign of an infinity, as does an abstraction.\(^2\) Represented by the analogical archetype and symbol, still the abstraction is understood independently of the analogical archetype or symbol. All would be immediate if not so, and there would be no abstraction. Whether or not abstraction is understandable independently of a representation, it still exists independently of representation. Required for understanding, abstraction is not representation of sensation, but awareness of sense.

Imagining a world whose properties are unimaginable being possible, indeed an infinity of such worlds, shows imagination can be a non-representational phenomenal experience. An example is the quantum in physics. When such, imagination is expressed as a “sense,” distinguished from a representational “sensation.” It might be thought a representation of what is imagined, but a representation is at best an analogical archetype, not the thing imagined. This latter is an abstraction, the set of all things like the analogical archetype. As such, it cannot be experienced in its entirety at any moment, only sensed at any particular time.

Having a sense without a representative sensation is possible. Sensation is immediate awareness. There being no immediate awareness of abstraction, there is no sensation of abstraction. There being awareness of abstraction, it is by another means. This is by “\(^1\)sense 3: b: a definite but often vague awareness or impression,”\(^3\) which is to “\(^2\)sense GRASP, COMPREHEND.”\(^4\) Manifest is “intuition c: the power or faculty of attaining to direct knowledge or cognition without evident rational thought and inference.”\(^5\) It is a primitive act of, “apprehension 1 a: the act or power of perceiving or comprehending.”\(^6\) As so, although a “vague awareness or impression,” a sense of understanding is still to “comprehend 2: to contain or hold within a total scope, significance, or

---

\(^1\)As Robert Audi indicates, 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. Audi 91.

\(^2\)An example of this occurs in number theory where, “The set-theorist Ernst Zermelo proposed that the number 0 is the empty set (\(\phi\)) and for each number \(n\), the successor of \(n\) is the singleton of \(n\), so that 1 is \(\{\phi\}\), 2 is \(\{\{\phi\}\}\), 3 is \(\{\{\{\phi\}\}\}\), etc. So every number except 0 has exactly one member.” [Stewart Shapiro, Thinking about mathematics: The philosophy of mathematics (Great Clarendon Street, Oxford, UK: Oxford University Press, 2000) 265.]

\(^3\)“sense” in Merriam Webster’s Collegiate Dictionary, tenth edition, eds. Frederick C. Mish, et. al. (Springfield, Massachusetts: Merriam-Webster, Incorporated, 1996) 1066.

\(^4\)“sense,” in Mish, et. al., 1066.

\(^5\)“intuition c,” in Mish, et. al., 614.

\(^6\)“apprehension 1 a,” in Mish, et. al., 57.
Abstraction and the Interpretation of Quality

As so, it constitutes “comprehension 1 c: the capacity for understanding fully.”

Sense can be mistaken if not behaving as if having it. Such behavior constitutes a pattern of specified things related by specified logical transitions. Understanding is indicated in the ability to identify the criteria of membership and relationship by defining the pattern, or to identify the content of membership and relationship by following the pattern. In Wittgenstein’s phrase it is, “to go on.”

Necessity of mathematical proofs illustrates this. Contra the claim it can be known mathematical proofs are necessary in a field, it can be replied since there is no awareness of all possible proofs in a field, it cannot be known all proofs are necessary in it. To the reply although not experiencing all proofs in a mathematical field, one has a sense all proofs are necessary, it can be replied only experience provides knowledge. Accepting this, since not experiencing all proofs in a mathematical field, how is it known a proof need not be necessary? Abstraction vanishes if a sense of truth is not knowledge, and the claim of the possibility of an unnecessary proof in a mathematical field is an abstraction.

Indeed all is abstraction. Experience occurring in time, and time being reducible to a nonexistent infinitesimal moment, experience is inherently abstraction of past and perhaps future moments not currently experienced. Since abstraction is unknowable by sensation, how is it knowable at all? And since all is abstraction, how is anything knowable at all? Not an object observed in itself, for instance, how is time knowable?

Resolution is by understanding knowledge of the abstract as a presumption or propositional attitude. Being primitive, whether it is a definition or sensation or something else is indeterminable. Such is beyond human knowledge. What is knowable, it is an a priori conception of truth. Knowledge must be a priori, and can be a posteriori. If the latter, what is considered and how it is considered is necessarily determined by an a priori preconception of truth or what is. It is the a priori which is known as a sense, and the a posteriori when judged by the a priori truth of relevance.

Abstraction is manifest in quality. Only existing in time and/or space, quality is irrational because time and space are extended. Thus, they are divisible into segments, and the segments are divisible into segments, and so on limitlessly. Limitlessly divisible, time and space are irrational, any calculation containing either or both having no product. Abstraction is simple identity. As

1“comprehend 2,” in Mish, et. al., 237.
2“comprehension 1 c,” in Mish, et. al., 237.
3Wittgenstein, Philosophical Investigations, 151, 59\textsuperscript{e}. See also 59\textsuperscript{e}-62\textsuperscript{e}, 80\textsuperscript{e}-84\textsuperscript{e}.
5“propositional attitudes. . . . are propositions asserting that something is believed, doubted, desired, and so on, in so far as such propositions are known independently of inference.” Bertrand Russell, An Inquiry into Meaning and Truth (Baltimore, Maryland: Penguin Books, 1962) 155.
such, it is unextended. Being unextended, it is indivisible. Because indivisible, abstraction is rational.

Sequence

Abstraction being identified by simple identity, and not existing in time, how is the same abstraction identified in time? No longer applicable is D. W. Hamlyn’s criterion, “If anything provides the principle of individuation for substances, or at any rate material substances, spatio-temporal history does,” because abstraction has no spatio-temporal history.¹ Even Hamlyn acknowledges this when qualifying, “at any rate material substances.” But if spatio-temporal history is the “principle of individuation for [material] substances,” yet cannot individuate abstract substances, how can the same abstract substance be individuated?

Indicating the same abstract substance can be individuated is the expression “this is the same idea I had before” is meaningfully assertable. Such individuation is explainable only by simple reidentification, this as primitive as initial identification. It is simply known to be the same abstraction. Even the current idea only being like the former idea, not being the same idea, requires recall of what is identifiable as the former idea, reintroducing historic identification of what has no spatio-temporal history. Otherwise abstract thought is constantly de novo.

Still this is incomplete, individuation of abstractions not confined to simple identity. Abstraction is distinguished not only in this way, but also by the sequences within which it can occur. Abstract mind can be sequenced with some things, and abstract body with other things. And different minds and bodies can be sequenced with different things. This sequencing function not only distinguishes abstractions, it also relates abstractions which can occur interchangeably in the same sequences.

Being self-identifying, abstraction is self-contained. Being self-contained, abstraction can be related to other elements, abstract or qualitative, only by sequencing. Abstraction being limited, an abstract sequence is limited, rendering the sequence rational. Quality being limitless, abstraction brings a limit to a qualitative sequence, rendering the sequence rational.

Abstract sequence is logically certain. A logic is identification of particulars. Such is intentionality, which is sequencing of particulars, which is coterminous awareness of particulars. Awareness being emergent, coterminous awareness of particulars is primitive. There are different instances of coterminous awareness of particulars, constituting different logics.

Whether there is intervening content between particulars determines how transition from one to another occurs. This is a function of content being qualitative or abstract, qualities always having intervening content and abstractions not. Qualities having intervening content, qualitative logic is

limitless. Suffering Zeno’s Paradox, it is uncertain. Abstractions not having intervening content, abstract logic is limited. Not suffering Zeno’s Paradox, it is certain.

**Worlds**

Truth can exist within an abstract world because an abstract world can be rational. Truth cannot exist within a qualitative world because quality is irrational. Only within a limited set is knowledge possible because an unlimited set is irrational. An irrational set has no product, when the product of a set constitutes knowledge. Knowledge is the product of a calculation, and the product of a calculation is identity of a composite. Even self-identity is a calculation, parts identified as whole and whole identified as parts. An unlimited set has no product, so there is no knowledge within it. Being such, an unlimited set is irrational.

Edmund Gettier’s counterexamples are irrational because they are outside the limits of the set, their irrationality proven by the possibility of any Gettier counterexample having its own Gettier counterexample.¹ This possibility being unlimited and unpredictable, and an unlimited unpredictable set defining irrationality, Gettier counterexamples are irrational. Science assumes all understanding occurs within a limit. Unlimited, knowledge is impossible. Rational activity is moving toward the limit.

There are two kinds of rational worlds. Either a world is supervenient with the properties of every complex within it deductible from the properties of its constituents, or emergent with the properties of at least some complex within it not deductible from the properties of its constituents. An emergent world might be thought irrational because incoherent, but is not. Any order is rational, whether coherent or not, even the wholly emergent world of chaos exhibiting predictable disorder.

An irrational world is a world in which not everything occurring within it is deductible from a constitutive principle or principles. In a world of one principle, all occurrences are deductible from the principle. In a world of more than one principle, all occurrences are deductible from one of the principles. All is coherent if no occurrence follows from more than one principle. If an occurrence does, it is ambiguous and the principles from which it follows are vague. Now a world is incoherent, and irrational because it is made coherent only by arbitrary choice.

Coherent and incoherent worlds are asymmetric. A coherent world need be determinate, and an incoherent world need not be determinate. Indeterminacy is a product of ignorance or error in a coherent world, and a product of freedom in an incoherent world. Order is explained by the supervenient structure encompassed in one principle of a coherent world. Disorder is explained by the emergent structure encompassed in more than one principle of an incoherent world. A rational world is constituted by one principle determining that world.

A nominal world is constituted by more than one principle determining that world. A chaotic world constituted by no principle cannot exist.

**Epistemology**

Analogue identity is likeness to a description or an archetype, Bertrand Russell’s conception of a propositional function illustrating descriptive likeness. Conformity to a rule being a relation, though, a proof is incomplete without conformity to a second rule identifying conformity with the first rule. And conformity with the second rule being a relation, a proof is incomplete without conformity to a third rule identifying conformity with the second rule, and so on, a proof being endless.

Relation is an unbroken path between two elements within a domain. If a broken path, how is an element prior to the break known to be the same element subsequent to the break?\(^1\) Elements in different domains are proven

---

\(^1\)Providing a proof of the equality of the angles of a triangle Ian Stewart avoids the pitfalls of Zeno’s paradoxes, of which he cautions, with, “Take hold of your cardboard triangle, turn it over, and put it back where it came from . . . . since it makes no difference where it goes in between, we do not need to talk about where it goes in between, and as a result we need not assume that it goes anywhere.” [Ian Stewart, *Concepts of Modern Mathematics* (New York: Dover Publications, Inc., 1995) 13.] Recursion is introduced to identify the triangle as the same whereby,

What we do need to know is where each point of the triangle finishes up. To do this we must have a way of labelling the points of the triangle, and the easiest way is to label all the points of the plane once and for all, so that we do not need to do everything all over again for a new diagram. [Stewart 13-14.]

Unstated is how it is known labeled “points of the triangle” before and after having turned “it over, and put it back where it came from” are the same “since it makes no difference where [they go] in between.” Not tracing an unbroken continuum of the points before and after having been turned over, there is no necessity in a point identified before “changing right to left and left to right” is the same point after “changing right to left and left to right.” [Stewart 11.] Further, it cannot be argued “new angle \(\triangle ABC\) is lying on top of old angle \(\triangle ABC\), so the two are equal” without reintroducing the same conundrum. [Stewart 11] How are the “new angle \(\triangle ABC\)” and “old angle \(\triangle ABC\)” known to be the “same” “since it makes no difference where [they go] in between?” A solution is identity as self-evident. Accepting William James’ “stream of consciousness” as “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,” [William James, “The Stream of Thought,” *Principles of Psychology*, in *Pragmatism: The Classic Writings*, ed. H. Standish Thayer (New York: Mentor, 1970) 142.] then any quality can uninterruptedly transmute into any other quality without exhibiting a concealed contradiction. [Ernest Nagel and James R. Newman, *Gödel’s Proof* (New York: New York University Press, 1986) 23.] Acknowledging this, all constituents of an unbroken continuum between any two recursive constituents are incorporated in endless reduction. Resolved is a difficulty introduced by E. J. Lowe whereby, “resemblance between particulars is always resemblance in some respect—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.” [Lowe 356.] Appropriate specification constituting a rule, however, how does a particular “in the relevant resemblance class . . . resemble the paradigm?” Paradigm being words and particular not, identity is not self-evident. Identity is
related by tracing an unbroken path between them, incorporating both into a
common domain. Proof is tracing such an unbroken path, mathematically
constituting identifying a dense st.\(^1\) It is material when physical, an unbroken
path of matter between limits. It is mental when phenomenal or conceptual.

Phenomenal it is semiotic, an unbroken path of consciousness between
limits.\(^2\) Conceptual it is logical, an unbroken path of elements between limits.
Semiotic is particular, not general; logic is general, not particular. Particular is
neither an analogical archetype nor analogical identity. General is either an
analogical archetype or analogical identity.

Experience composes qualitative and abstract elements. Qualitative
experience is sensation, and abstract experience is sense. Sensation is extended
in time and/or space, and sense is extended in neither. Grouped and ungrouped
qualitative elements, whether sensate or imaginative, are indistinguishable.
Therefore, distinguishing grouped and ungrouped qualitative elements is not
qualitative. By elimination, it is abstraction.

Observationally abstraction is intentional. It can appear an unextended sense
of self-identity, or an extended sense of self-identity containing or contained
within another identity. There is no sensation of a quality distinguishing it as
fused or diffused elements. There is no constant sense of a quality
differentiating it as fused or diffused elements. There is a sometimes sense of
something differentiating it as fused or diffused elements. Therefore inconstant
abstraction distinguishes quality as whole or parts.

Because inconstant, sense of abstraction can be mistaken, error determined
by with what other quality or qualities a sensed quality is analogical identified.
Sense as fused or diffused elements is sense of analogical identity of whole
with whole or part with part. Identity of elements with what is self-evidently
whole distinguishes elements as fused; identity of an element of elements with
what is self-evidently part distinguishes elements as diffused. Since what is
self-evidently part can be mistaken in its own identity, error is determined by
with what other quality or qualities this sensed quality is analogical identified,
etc.

\(^1\)“An ordered set is said to be dense, if it contains at least two elements and no neighboring
elements. A dense set is always infinite, because every finite set containing at least two
elements has also neighboring elements.” [E. Kamke, *Theory of Sets*, trans. Frederick

\(^2\)William James’ “stream of consciousness” manifests this, being “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.” [William James, “The Stream of Thought,” *Principles of
Psychology*, in *Pragmatism: The Classic Writings*, ed. H. Standish Thayer (New York: Mentor,
1970) 142.]
Cosmology

Shaped by the Mylesians, Classical Greek philosophy identifies being 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’).”\(^1\) Supposing “being qua being” provides uniformity making possible “an algorithm, a mechanical proof procedure, that . . . might exhaustively describe or ‘rationally reconstruct’ . . . all rationality worthy of the name.”\(^2\) All is determined, providing the necessary order of “. . . ‘scientism’— an exaggerated and ideologically explainable respect for a certain mistaken image of science.”\(^3\)

Despite qualitative ambiguity, Charles Saunders Peirce drew the conclusion that all cognitions are the results of inference, and so are logically determined by prior cognitions of the same object. The fact that this involves an infinite regress Peirce accepted once he was convinced that the regress was not vicious.\(^4\)

But the regress is vicious.

Just as element can be conjoined with element to generate such “an infinite regress,” the regress itself can be conjoined with its contradictory, \(R\) and not \(R\), denying every constituent of Peirce’s regress tit-for-tat, \(r\) subscript 1 and not \(r\) subscript 1, \(r\) subscript 2 and not \(r\) subscript 2, . . . , rendering Peirce’s regress vicious because \(r\) subscript n and not \(r\) subscript n satisfies the condition of membership in \(R\), which is being \(r\). Eliminating this possibility requires a new rule, which can be conjoined with its contradictory, \(\{R\}\) and \(\{\text{not } R\}\), etc., rendering any “regress involved in predication” vicious.\(^5\)

---

\(^5\)In recognition Ernest Nagel and James R. Newman observe, “Non-finite models, necessary for the interpretation of most postulate systems of mathematical significance, can be described only in general terms; and we cannot conclude as a matter of course that the descriptions are free from concealed contradictions.” [Ernest Nagel and James R. Newman, *Gödel’s Proof* (New York: New York University Press, 1986) 22-23.] Similarly Stephen Barker queries, “Could there be any way of establishing that the revised deductive systems for set theory and number theory . . . were free from inconsistency and would not eventually give rise to contradictions also?” [Stephen F. Barker, *Philosophy of Mathematics* (Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1964) 91-92.] Realized is the entropy principle introduced by Rudolf Clausius whereby, “The energie of the universe is constant while entropy [randomness] strives toward maximization.” [James Gilbert, *Redeeming Culture: American Religion in an Age of Science* (Chicago: University of Chicago Press, 1997) 22.] In the extreme, “All change will cease. This prediction, called the heat death of the universe, has been much discussed by philosophers. The tendency toward this final state would seem an inevitable consequence of the second law of thermodynamics, although it would lie very far in the future.” [Douglas
Undaunted, Peirce responds,

"[a world in which induction would fail as often as lead to truth] would not be disorder, but the simplest order; it would not be unintelligible, but, on the contrary, everything conceivable would be found in it with equal frequency." [1]

Presuming "no society has ever known its language to be anything other than something inherited from previous generations, which it has no choice but to accept," [2] argument proceeds by conjoining element to element in an infinitely expansive set from which every constituent necessarily follows. [3]

Still, contradiction endlessly supplants qualification and qualification endlessly supplants contradiction, interminably devolving toward the unique identity of casuistry. Eternally approached, "an algorithm, a mechanical proof procedure, that . . . . might exhaustively describe or 'rationally reconstruct' . . . all rationality worthy of the name" is unattainable. Neither can such a "proof procedure . . . 'rationally reconstruct' . . . all rationality worthy of the name" because contradiction is resolvable differently, generating internally consistent alternative algorithms or mechanical proof procedures. Scientism fails, the universe never "free from concealed contradictions." [4]

**Ontology**

Relevantly, "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." [5] Scientism overlooks problematic character of "categories." Anaximander identifies being as content (*apeiron*); Anaximenes identifies being as content (*apeiron*) and form (condensation and rarefaction). Since same content can appear as whole (condensed) or parts (rarefied), form is not

---

3. Because axiomatic theory presupposes, "that anything at all can play the role of the undefined primitives of points, lines, planes, and so on, so long as the axioms are satisfied," any element can be conjoined with any other element in a set from which the element necessarily follows. [Stewart Shapiro, *Thinking about mathematics: The philosophy of mathematics* (New York: Oxford University Press, 2000) 151.] This conjunction can be conjoined with any other element in a set from which the conjunction necessarily follows, and so on infinitely. Each conjunction meeting the condition of an axiomatic system, and an axiomatic system meeting the condition of a language, "That is why the question of the origins of language does not have the importance generally attributed to it." [Saussure 72.]
content. Not constituting content, it is ontologically distinct, “being qua being” now complex, not simple.

Simple are content and form. Content composes quality; form composes abstraction. Quality is intrinsically temporal and spatial, identified as different in different time and space. Abstraction is accidentally temporal and spatial, identified as same in different time and space. Complication occurs because limitlessly divisible in time and space, whether quality is one or many is ambiguous.\(^1\) Cardinality of subsets of a nonempty set being greater than cardinality of members of the set exhibits this. Subsets identifying forms of set content, content does not identify form.

Unobservable like theoretical entities, unlike theoretical entities which are substantial in some sense, like physical forces abstraction is insubstantial. Also like physical forces abstraction infuses quality without measurably affecting it.\(^2\) Both being “something,” they constitute impalpable entities ontologically distinct from quality.\(^3\) Not substituting for quality, abstraction does not transmute from quality as Colin McGinn presumes.\(^4\) Abstraction is the fugitive “meaning” identifying a symbol.\(^5\) That physical forces can vary in time and space renders them physical. That abstraction does not vary in time and space renders it metaphysical.

Intrinsically indeterminate, as determined when determination is observationally indistinguishable presumes an infusive non-qualitative identifier. Abstraction infuses quality as do physical forces, though parasitic to particulars when physical forces are universal. Still, both put the lie to, ““Two

---

\(^1\)By illustration, but what are the simple constituent parts of which reality is composed?—What are the simple constituent parts of a chair?—The bits of wood of which it is made? Or the molecules, or the atoms?—“Simple” means: not composite. And here the point is: in what sense ‘composite’? It makes no sense at all to speak absolutely of the ‘simple parts of a chair’. Wittgenstein, nt. 47, 21\(^e\).

\(^2\)Identifying the extent of a quality, an abstraction recursively identifies its expanse and iteratively identifies its limit at every level of scale, permeating the quality in this sense.


\(^4\)The nature of consciousness is a mystery in the sense that it is beyond human powers of theory construction, yet there is no sense in which it is inherently miraculous. . . . Epistemologically, consciousness outruns what we can comprehend, given the ways our cognitive systems are structured . . . . Ontologically, however, nothing can be inferred from this about the naturalness or otherwise of the object of our ignorance: what cannot be known about is not thereby supernatural in itself. Colin McGinn, The Character of Mind: An Introduction to the Philosophy of Mind (Oxford, England: Oxford University Press, 1997) 42. The reason that consciousness must have a hidden structure, intrinsic and essential to it, is that the unknown properties that link it to the brain must be internal to consciousness. Something about the nature of consciousness itself must explain how it is possible for it to emerge from matter; the explanatory principles cannot lie outside of consciousness. This is not the old idea of the unconscious, a mental system that exists side by side with the conscious mind; rather, conscious states themselves have an aspect they do not present to introspection. McGinn 44.

\(^5\)Syntactical meaning is incomplete because syntax is a sequence when sequence is ambiguous, identifiable cardinaly or ordinally. Sequence is a matter of interpretation, and interpretation is abstract identity.

238
different objects cannot occupy the same place at the same time.”

Abstraction infuses physical and phenomenological quality when physical forces infuse only physical quality, although appearing to infuse phenomenological qualities as in dreams despite not being present. Constituted is normativity, resolution of qualitative indeterminacy.

Conclusion

Contrary to expectation, the practical requires abstraction because truth is limited to an abstract world. Practicality being the ability to predictably alter experience, this is possible only with abstract elements. Prediction is a calculation identifying the product of elements in specified sequential consideration. Endlessly divisible in endless variation, qualitative elements “can be expressed as an infinite decimal with no set of consecutive digits repeating itself indefinitely” of an irrational number. Being irrational, a calculus of qualitative elements has no product. Being indivisible, abstract elements are limited and rational as such. Being rational, a calculus of abstract elements has a product. Only when calculating the product of abstract elements is predictive calculation of the practical possible.


2Being one or many, when which is observationally indeterminable, quality is intrinsically ambiguous. Resolution of ambiguity is normativity. There being no ambiguity, there is no normative. Normativity arises from the character of quality. Occurring in time and/or space, quality presents distinct particulars forming an indistinct particular. Abstraction occurs in neither time nor space, presenting distinct particulars not forming an indistinct particular. Being particulars forming a particular, qualities are sequentially transitory, not initiatory. Being particulars not forming a particular, abstractions are sequentially initiatory, not transitory.

3“irrational number,” in Mish, et. al., 619.
CHAPTER TWENTY THREE

Hegel’s Early Response to Pyrrhonian Skepticism

Ioannis Trisokkas

In his early essay *The Relation of Skepticism to Philosophy* (RSP),1 Hegel provides an extended discussion of Pyrrhonian Skepticism (PS). He argues that besides ‘dogmatism’ and ‘skepticism’ there is also ‘philosophy’, a term which he uses interchangeably with ‘speculative philosophy’. This denotes the standpoint of reason, which investigates and produces knowledge of a certain dimension of what-there-is: the rational.

Hegel aims to defend the thesis that PS has a ‘noble essence’ that (a) works successfully against dogmatism, but (b) fails against speculative philosophy. Peculiarly, it owes its failure partially to the fact that speculative philosophy ‘incorporates’ it, that it is ‘in its inmost heart at one with […] true philosophy.’

The other factor related to this failure is the element of ‘unification’ with which the ‘incorporation’ of the ‘noble essence’ of PS is closely linked.

My paper aims to examine RSP and evaluate its effectiveness against PS.

Pyrrhonian Skepticism

Scientific inquiry takes place by means of positing judgments in the universe of discourse. These judgments emerge in this universe with the intention of presenting themselves as true – they are, namely, truth-claims. It is a fundamental trait of these claims that immediacy is a necessary part of their existence. This is so because when a truth-claim appears for the first time, its relation to other truth-claims is not made explicit. Due to this moment of immediacy, there is a necessary state of affairs which designates the truth-claims’ having an equal truth-status with their negations. This must be so, for negative truth-claims also appear immediately in the universe of discourse and in a context of immediacy no privileging of a truth-claim over its negation can be made; for in this context its truth emerges only from its sheer presence, from its sheer intention to be true.

---


2RSP 227.
In this way there reigns a state of equipollence between conflicting truth-claims. The thrust of PS is that it ‘manifests’ that such conflicts can never be resolved and that, therefore, science can never achieve knowledge of truth. This ‘manifestation’ is made in terms of the so-called argument from the criterion of truth. The Pyrrhonist argues that the scientists’ attempt to resolve a conflict by privileging one of the truth-claims is doomed to failure. For, in order to achieve this ‘privileging’ the scientists must explicitly say why the one rather than the other truth-claim is true; they must, in other words, specify a criterion of truth which will ground the one’s superiority over the other. Yet, in doing so they have to say something further about the real, for a conflict about the truth of the real can be resolved only if more is said about it. Accordingly, the criterion is itself a truth-claim, and since, like all truth-claims, it must exhibit a moment of immediacy, it finds itself in conflict with its negation. In this way, vicious infinite regress or circularity soon ensues. Given that these phenomena have a formal character in the argument, their presence establishes the impossibility of ever resolving a conflict in the universe of discourse.1

Skepticism, Dogmatism and Speculative Philosophy

In the RSP Hegel suggests that the refutation of PS could be accomplished through a theory which would show that the cognitive standpoint of reason can provide us with knowledge of a specific dimension of what-there-is: the rational. The truth-claims posited in the universe of discourse claiming knowledge of the rational are called truth-claims-of-reason. The cognitive standpoint of reason is exemplified by ‘philosophy’, which therefore aspires to produce knowledge of the rational. Yet, within the general sphere of ‘philosophy’ one distinguishes between two ‘attitudes’ towards the manner in which this knowledge is obtained: (a) ‘dogmatism’ and (b) ‘speculative philosophy’. Hegel illuminates this difference by examining the two attitudes’ relation to PS, and in particular their relation to the so-called Five Modes of Agrippa. The significance of Hegel’s discussion of the Five Modes is that it distinguishes which of the ‘principles’ present in the universe of rational discourse are dispensable (in the sense that they fall prey to PS) and which are not. The term ‘dogmatism’ will here denote exactly that ‘attitude’ in the sphere of ‘philosophy’, which proceeds by positing truth-claims from the cognitive standpoint of reason, but whose approach to knowledge-acquisition is fundamentally determined by those dispensable ‘principles.’ Hegel correctly points out that the Five Modes were meant to refute not simply ‘experience’ and ‘sense-perception’, but mainly dogmatism and speculative philosophy.2 But, he argues, while the Five Modes succeed in refuting dogmatism, they are ‘completely ineffective against philosophy.’3 The

---

2 RSP 243, 244-245, 240.
3 RSP 244-245.
general reason given for this diversity is that ‘they contain reflective concepts.’

He explains:

Directed against dogmatism they appear from the side where they belong to reason, setting the other term of the necessary antinomy alongside the one asserted by dogmatism; directed against philosophy, on the other hand, they appear from the side where they belong to reflection. Against dogmatism they must necessarily be victorious therefore; but in the face of philosophy, they fall apart internally, or they are themselves dogmatic.

The concluding status of the last sentence is not apparent to us. What seems clear is only (a) that the Five Modes can thwart dogmatism because of their ability to give rise to ‘antinomies’ and (b) that they cannot destroy philosophy because when they confront it ‘they appear from the side where they belong to reflection.’

The Refutation of Dogmatism

Hegel sees the Five Modes as providing the various parts of a single argument. The formulation of this argument will first help us understand what the essence of dogmatism is and exactly why it cannot escape PS. Based on this result, we will then be able to define the essence of speculative philosophy.

The Hegelian-Pyrrhonian argument from the Five Modes against dogmatism can be described as follows. The dogmatist holds on to a certain content and treats it as if it were the absolute truth: its truth arises solely from itself, independently of any relations to other contents. Reason, however, finds itself unable to determine the posited content unless it relates it to those contents which are opposed to it. This realization of reason exemplifies the Third Mode (‘from relation’) and discloses the limitedness of the posited content.

The dogmatist conceives such relation as a sign of the vanishing of truth and wishes to destroy it. He, therefore, posits a content as the ground of the truth of one of the relata, which immediately obtains the status of what-is-grounded. It might be the case that a content’s determination requires its relation to its opposite, but in terms of truth, a grounding content can easily effect their separation. This grounding-relation is meant to secure the absoluteness of the initially posited content. Reason, though, finds itself wondering how the grounding content would ground the initially posited content and not its opposite. This question brings forth three responses; their failure leads to the collapse of the relation between truth and ground.

1 RSP 245.
2 RSP 245.
3 RSP 245-246.
(a) **Circularity**: The first response asserts that the grounding content grounds a content A and not a content B because it owes its existence to A: the grounding content comes into being because of A, not because of B. But this response entails that reason finds itself unable to determine the grounding-relation unless both of its relata are assigned the status of ground. This realization of reason exemplifies the Fifth Mode (‘from reciprocal dependence’) and discloses the circularity of the grounding-relation. Yet, the dogmatist himself finds circularity and truth incompatible and, therefore, wishes to destroy the former, without, however, destroying the grounding-relation.

(b) **Hypothesis**: To achieve this, the dogmatist asserts that the grounding content will be taken as being the self-grounded or ungrounded ground of A by hypothesis. Reason, however, acknowledges that, if the assignment of the ground-status is made only by hypothesis, then two irresolvable problems arise: (a) the truth of content B could be said to be equally validly grounded by a self-grounded ground; and (b) since the grounding content is simply assumed to be true, its own opposite content could equally validly be posited as true. This realization of reason is exemplified by the Fourth Mode (‘from hypothesis’) and is directed against the arbitrariness of the grounding-relation. Nonetheless, it is not only us, but also the dogmatist himself, who finds this arbitrariness incompatible with his idea of science and truth and wishes to move beyond it; he still insists, though, in maintaining the grounding-relation between contents as the locus of truth.

(c) **Infinity**: Aiming to achieve this, the dogmatist argues there is a reason why the grounding content grounds the truth of A and not of B or why it itself is the truth and not its negation: There is an objective ground that grounds the grounding content. But reason asks itself whether there is a reason why that ground is objective and not subjective. Since it cannot posit its objective status by hypothesis, it posits another content as the objective ground that grounds another ground. Clearly, an infinite regress of grounds now ensues. This exemplifies the Second Mode (‘from infinite regress’), which represents the final stage of the dogmatist’s futile appeal to the notion of ground.\(^1\)

The failed appeal to the grounding-relation forces the dogmatist to return to the simple relation between a content and its opposite. What he desires is the absolutization of one of those contents. Yet, he also believes that his desire can be fulfilled only through the grounding-relation; thus, he now finds himself in an impasse, since, firstly, the grounding-relation has proven unsuitable for this desire’s satisfaction and, secondly, what appears is a manifoldness of conflicting contents.

Instead of a singular content, the dogmatist confronts a universe whose essence is pervaded by a plurality of contents. This situation exemplifies the First Mode (‘from difference’), which discloses that no content can be said to be true unless one acknowledges that its relation to its opposite is part of that truth. But the dogmatist sees here nothing but a contradiction. To escape from

---

\(^1\)RSP 245.
Hegel’s Early Response to Pyrrhonian Skepticism

In this impasse, he relapses to his first position, the positing of an (allegedly) absolute content, by simply ignoring the presence of its opposite. In this way, the argument repeats itself endlessly, showing thereby how the dogmatist imprisons himself in it.

This argument reveals the essence of dogmatism, because it clarifies, firstly, what it aims at and, secondly, why it cannot achieve it. It aims at the resolution of conflict but admits an outcome as successful only if the one of the contents has totally vanished and the other has been absolutized. The essence of dogmatism is that it desires to privilege one part in a relation of opposition over the other and thereby absolutize it. It is because of this essence that the only means available to dogmatism for accomplishing its aim is through the notion of ground.

Dogmatism collapses, for the notion of ground allows one to manifest the impossibility of ever privileging one content over its opposite. The Five Modes, therefore, have been successful in their attack against dogmatism. Hence the same reason which the dogmatist employs in order to resolve conflicts manifests that they cannot be resolved if the aim is to privilege one content over another. What Hegel wants to emphasize here is reason’s inability to resolve a conflict by privileging one of its parts. But, Hegel immediately adds, this peculiar employment of reason as a means to resolve conflicts belongs to dogmatism, not to speculative philosophy.

As noted, conflict appears in the universe of discourse as soon as a negative truth-claim is posited therein. The aim is to resolve the conflict. The dogmatist maintains that conflicts between truth-claims of reason can be resolved if a ground is posited which would allow the privileging of the one truth-claim over the other. Yet, the argument from the criterion of truth has manifested that such privileging cannot be achieved. Hegel has made it clear that this failure to refute PS through grounding is intrinsically related to the posited aim, that is, the ‘privileging’ (or ‘absolutizing’) of one truth-claim of reason over another. This means that if one shows that conflicts in the realm of reason could be resolved without this privileging taking place, it would no more be necessary to employ the ‘reflective’ notion of ‘ground’ in such an enterprise and, therefore, the argument from the criterion of truth could not be formulated in the first place. This conclusion is extremely important, for it shows that it is because it is the ‘scientists’ themselves who attempt to resolve the conflicts by means of a ground that the Pyrrhonian argument is so effective against them.

Speculatibe Philosophy and the Rational

In contradistinction to dogmatism, speculative philosophy outlasts the Five Modes on the following general grounds:
[As] directed against reason [...] the Five Modes retain as their peculiarity the pure difference by which they are affected; their rational aspect is already in reason.¹

In this section I will try to decipher this statement by examining those passages which make specific suggestions as to how speculative reason escapes the challenge raised by the Five Modes.

Let us first see how speculative philosophy responds to the First Mode:

As far as the First Mode (“from difference”) is concerned, the rational is always and everywhere self-identical; pure inequality is possible only for the understanding [i.e. dogmatism], and everything unlike is posited by reason as one.²

Recall that the First Mode was directed against the dogmatist’s belief that a content can be posited as true in isolation from other contents: What is primordially present is not an absolute content but a multiplicity of conflicting contents. Yet, while this makes a point against dogmatism, it fails to undermine speculative philosophy; this is so because even though reason manifests that there is only one true content, the latter is not opposed to other contents. In the realm of reason, difference is a fundamental element of the self-identical structure of the rational. Speculative philosophy accepts neither the positing of an absolute content nor the simple presence of a multiplicity of contradictory contents; it accepts them both, in the sense that the multiplicity of contents is also posited as a singular content. Pure isolation is alien to speculative philosophy.

Speculative philosophy defies the Third Mode (‘from relation’) as well, because

it cannot be shown of the rational, in accordance with the Third Mode, that it only exists within the relation, that it stands in a necessary connection to an other; for it is itself nothing but the relation.³

The Third Mode undermined the dogmatist’s claim that a content can have a determinate existence without being related to its opposite; rather each content ‘exists only through and in’ its relation to its opposite.⁴ Thus, the dogmatist’s belief in an undifferentiated, absolute content collapses when confronted with this Mode. Yet, speculative philosophy does not face this problem, for the rational relates to no opposite, not because it can isolate itself from it and ignore it, but because it is the relation between opposites itself. Philosophy’s

---

¹RSP 246.
²RSP 246.
³RSP 246.
⁴RSP 245.
purpose, then, is not to posit and defend absolute contents, but to manifest positive (or ‘unifying’) relations between opposites.

It is now the turn of the Fifth Mode (‘from reciprocal dependence’) to prove ineffective against speculative philosophy:

Since the rational is the relation itself, it escapes from the circle (or the Fifth Mode: ‘from reciprocal dependence’), in contradistinction to the relata, which, when they are posited by the understanding [i.e. dogmatism], are supposed to ground one another, and therefore should necessarily fall therein; indeed, with respect to the relation itself nothing is reciprocally grounded.¹

Dogmatism has it that a content could be grounded upon a content that it itself brings into existence. The Fifth Mode manifests the absurdity of this claim by clarifying that in such grounding-relation both contents deserve the ground-status. Still, while this destroys dogmatism, it fails to make a point against speculative philosophy, because, since the rational is the relation itself, it has no other to which it could relate in terms of ground.

Speculative philosophy escapes from the Fourth Mode (‘from hypothesis’) as well, since

the rational is not an unproved assumption, in accordance with the Fourth Mode, so that its counterpart could with equal right be presupposed unproved in opposition to it; for the rational has no opposed counterpart; it includes both of the finite opposites, which are mutual counterparts, within itself.²

The Fourth Mode was directed against the ‘hypothetical’ positing of a ground, disclosing that this is no truer than the hypothetical grounding of its opposite. This notion of ‘hypothetical grounding’, however, since the rational has no opposite, is not relevant to speculative philosophy.

Finally, since speculative reason’s activity has been shown to be alien to the idea of positing absolute, undifferentiated contents and adjudicating between them by means of grounds, it comes as no surprise that speculative philosophy undermines the Second Mode (‘from infinite regress’) as well. Hegel writes:

The two preceding Modes both contain the concept of a ground and of a consequent, according to which, one term would be grounded through an other; since, for reason, there is no opposition of one term against another, these two Modes, as well as the demand for a ground that is posited in the sphere of oppositions and repeated endlessly (in the Second Mode: “from infinite regress”), become

¹RSP 246-247.
²RSP 247.
irrelevant. Neither that demand for a ground nor this infinity is of any concern to reason.¹

Indeed, if the ‘demand for a ground’ is not present in speculative philosophy, then obviously no infinite regress of grounds can be found therein and, therefore, the Second Mode proves to be inapplicable to speculative reason.

All in all, the Five Modes are ineffective against speculative philosophy because the elements upon which they base their attack against dogmatism are not ‘determinative’ of speculative philosophy. These elements, which exemplify the way in which the Five Modes ‘retain [...] the pure difference by which they are affected’, are (a) the tendency to privilege or ‘absolutize’ one of the relata in a conflict and (b) the tendency to achieve this through the positing of a ground. Lacking such ‘tendencies’, speculative philosophy has a completely different character from dogmatism. Its subject-matter, knowledge of which will manifest what-there-is-in-truth, is neither a content that relates to an other in terms of opposition, nor a content whose essence contains no opposition. It is the explication of the containment of opposition within a unitary element that distinguishes speculative philosophy from dogmatism.

Yet, how would this character of speculative philosophy enable the resolution of conflict between truth-claims-of-reason? To respond to this question we must clarify the relation between speculative philosophy and the ‘noble essence’ of PS.

Contradiction and Unification of Truth-Claims-Of-Reason

PS purports to show that the resolution of conflict between truth-claims is impossible. It has now been clarified, however, that the successful manifestation of this impossibility requires that the ‘scientists’ have two unwarranted beliefs: (a) that the conflict can be resolved solely by privileging one of the truth-claims involved and (b) that this privileging can be achieved by positing a ground. Speculative philosophy professes to have escaped the Pyrrhonian attack, which must mean that it has discovered a way to resolve a conflict without assenting to those two beliefs. Thus, if the Pyrrhonian challenge is that we should resolve the conflict between truth-claims-of-reason without privileging one of them, then we are indeed able to respond successfully to it.

Hegel’s insight is that a conflict can be resolved, not by privileging one of the relata, but by uniting or incorporating them in a ‘higher’ truth-claim-of-reason. This is the manifestation of the rational itself and, therefore, is not an absolute content that has nothing to do with opposition. The ‘content’ of the ‘higher’ truth-claim-of-reason is nothing but the relation between conflicting ‘lower’ truth-claims-of-reason. The ‘higher’ status of the third truth-claim has,
then, a simple meaning: it is just the truth-claim whose determinate content’s analysis would produce truth-claims-of-reason whose relation is one of conflict. Here is how Hegel puts it:

If in some proposition which expresses knowledge of the rational, what is considered is the reflected aspect of it – that is, the concepts which are contained in it – taken in isolation and the manner in which these concepts are connected, then what must show itself is that these concepts are at the same time sublated or that they are united in such a way that they contradict themselves; otherwise, such proposition would be not one of reason, but only of the understanding [i.e. dogmatism].

Also:

[…] Every such proposition of reason lets itself being dissolved into two absolutely conflicting [propositions] – e.g. God is cause and God is not cause […].

Clearly, there is a problem here: the exact relation between (a) the notion of contradiction and (b) the ‘higher’ truth-claim-of-reason that unites contradictories. For since it aspires to resolve conflicts, speculative philosophy must conceive such events as problematic, which means that it accepts the validity of the law of non-contradiction. Yet, this resolution’s formulation as the unification of contradictories seems to deny the law’s validity and affirm the truth of a contradictory state-of-affairs. Thus, there should be a concrete difference between (a) that state-of-affairs which exemplifies a problematic contradiction and (b) that one which exemplifies a non-problematic contradiction. A problematic contradiction is one that relates two (allegedly) absolute truth-claims. Consequently, this absoluteness must be what is missing from the non-problematic contradiction. Hegel writes:

The so-called law of non-contradiction is so far from possessing formal truth for reason, that on the contrary every proposition-of-reason must, with respect to the concepts, contain a violation of it; for a proposition to be merely formal, means this for reason: to be posited purely for itself, without the parallel affirmation of what is contradictorily opposed to it; and just for this reason it is false. To recognize the law of non-contradiction as formal means, then, to obtain at the same time knowledge of its falsity.

---

1RSP 229.
2RSP 230.
3RSP 230.
Thus, there is a way in which the law of non-contradiction is invalidated in the realm of reason and it is only through this event that knowledge of the rational is actualized.

Contradiction must, therefore, be present in the ‘higher’ truth-claim-of-reason, where the conflicting truth-claims are united, but not in such a way that these truth-claims would be posited as absolute. This specific response reveals that Hegel’s idea of speculative philosophy is one of an activity that moves forward by constructing positive unities of contradictory elements; in contrast, dogmatism has the image of an activity that moves backwards by constructing grounds. While dogmatism is regressive, speculative philosophy is progressive. Nevertheless, there is a feature which they both share: The motor behind both movements is the essence of Pyrrhonianism, the positing of contradictories. But whereas dogmatism regards contradiction as a vice that needs to be ostracized, speculative philosophy admits its unavoidability and seeks to do justice to it in its attempt to arrive at knowledge of truth. ‘True’ philosophy could not exist without the incorporation of Pyrrhonism’s ‘pure negativity’ or, as Hegel puts it,

true philosophy has necessarily a negative side, which [...] is directed [...] against all limited [i.e. posited as absolute] contents. ¹

It is exactly because of this incorporation of ‘negativity’ within philosophy that Hegel characterizes the essence of PS as noble² and claims that ‘[Pyrrhonian] skepticism itself belongs to philosophy.’³ This also reveals in what sense ‘the rational aspect of the Five Modes is already in reason’; for like philosophy, they too depict all (allegedly) absolute contents as being limited by their opposites. (Such relation of opposition is, for Hegel, also a ‘reflective’ one; this is why he says that the Five Modes ‘contain reflective concepts’). They must, therefore, fail against philosophy because what they do not deny, it also does not deny.

The Pyrrhonist is right when he affirms the presence of contradictory relations; yet, his manifestation of the impossibility of resolving these contradictions does not work against speculative reason. Hegel believes that the unification of contradictories is immune from the attack of Pyrrhonism. He conceives this unification in such a way that it does not involve the elimination of contradiction altogether. It eliminates only a certain dimension of the phenomenon of contradiction, the problematic one; this is the dimension that exemplifies the one contradictory’s indifference against the other; this means that the problematic contradiction allows or even supports the possibility that the contradictories’ relation is resolvable in favour of the one of them. This is why Hegel believes that the ‘formal’ aspect of the law of non-contradiction is present here and that through this aspect it is false. For this reason, the

¹RSP 227-228. See also RSP 229, 230, 237.
²RSP 249-250.
³RSP 234, 237.
dogmatists approach knowledge-acquisition the way they do because they cherish only the ‘formal’ aspect of this law.

This does not mean that Hegel denies the validity of the law altogether; therefore, he does agree with those scientists and the Pyrrhonists that contradictory relations must be resolved. Yet, this resolution involves the maintaining of the contradictory relation in a special unity expressed by the ‘proposition of reason’ (Vernunftssatz). It is this paradox and this notion of the Vernunftssatz we need to clarify if we want to understand how exactly Hegel resolves the Pyrrhonian problematic. The realization of this task, however, must wait for another paper, since it cannot be done from within the textual framework provided by RSP.

Conclusion

In RSP Hegel proposes to resolve the Pyrrhonian problematic through abandoning the idea of grounding an absolute truth-claim and introducing the idea of unifying contradictory truth-claims in a ‘higher’ truth-claim-of-reason. The notion of a ‘higher’ truth-claim, however, requires further explication. All one knows right now is that it can be analyzed into a group of contradictory judgments, the ones which it unites. Yet, its simple characterization as the unity of contradictories says nothing to help us distinguish it from that state-of-affairs which denotes a simple conjunction of conflicting truth-claims. True, the Pyrrhonist can no longer manifest that such conjunction results necessarily in the abandonment of scientific inquiry, but this has now been conditioned upon the ability of speculative reason to prove that a unity of conflicting truth-claims can occur which is irreducible to their simple conjunction. Otherwise, all Hegel has done is to reinstate the problem itself.

Thus, the ‘higher’ truth-claim-of-reason, despite being nothing but the ‘unity’ of contradictories, has a content that is qualitatively different from their simple conjunction. But how exactly does a content like this emerge? What form would it have? Would it still be possible to accommodate it in a judgment or should one start invoking other forms of truth-claim? And how would one do that if what appears immediately in the universe of discourse is only the judgment? Is it not the case that for any posited judgment its negation immediately pops up next to it? How would then this infinite progression fate better than the infinite regression of the dogmatist? To all these questions RSP offers no response.

What we should be looking for is the meaning and structure of what Hegel calls Vernunftssatz, and, more specifically, the meaning and structure of the ‘higher’ Vernunftssatz. We need to discover how it is possible to move from a truth-claim-of-reason whose structure as a judgment forces it to be conjoined with its negation to a truth-claim-of-reason whose structure allows it to accommodate and express that conjunction in a non-problematic, truthful manner. Such theory of the Vernunftssatz, however, can no longer be given in terms of a dialectical confrontation with Pyrrhonism. For the Pyrrhonist expects
that the speculative philosopher will now posit truth-claims in the universe of discourse, stating what he believes to be the true nature of the Vernunftssatz. But, the Pyrrhonist would argue, whatever it is that he would like to say, he would not be able to say it, since for any truth-claim he makes, its negation will immediately pop up in that universe.

In terms of the resolution of the Pyrrhonian problematic, what the speculative philosopher has achieved in the context of the RSP is only to have pointed out to the Pyrrhonist that, given his own description of the problem, one could attempt to resolve a conflict not only by means of a ground, but also by means of unification. The crucial moment has now been reached when the Pyrrhonist simply waits to hear what the positive characterization of that possibility is. Therefore, the immediate concern of the speculative philosopher is to provide this characterization without simply reinstating the Pyrrhonian problematic. The result of such characterization should be a full-blooded theory of the Vernunftssatz, an account of how it is possible to move from that structure of the judgment which 'provokes' conflicts to a structure that transforms those conflicts into something positive and truthful. My thesis, which, due to space limitations, I cannot defend here, is that such a theory is indeed offered by Hegel in the third part of his Science of Logic (Logic of the Concept).
according to Michael Dummett, the meaning of an assertion is given by its justification: justification completely exhausts meaning. In this perspective, semantics and epistemology are strictly entangled: knowing the meaning of a sentence amounts to knowing the justification one must offer for it. Contemporary contextualist epistemologists have convincingly shown how the notion of justification can be contextualized: justification depends on a specific issue-context. If we combine Dummett's epistemic account of meaning with a contextualist perspective on justification, we obtain an outline of a contextualist theory of meaning. This theory is undoubtedly new and original, but it suffers from some serious drawbacks.

Our paper is structured as follows. In section 2 an epistemic account of meaning is sketched, based on Wittgenstein's slogan "meaning is use" and Dummett’s theory of meaning as justification. In section 3, we show how a specific notion of justification (the one proposed by foundationalism) can be contextualized. S may be justified in uttering \( p \) in context \( C_1 \), but not justified in uttering \( p \) in context \( C_2 \): justification depends on a specific issue-context, which determines the appropriate objector-group. Precisely because of this reference to appropriate objector-groups, the proposal faces a difficulty that we try to overcome. In section 4, we outline a seemingly attractive way of contextualizing meaning, and we point out its major shortcomings. While in Dummett's semantic theory the meaning of an assertion is given by its justification, in epistemology a theory of justification aims to clarify the meaning of the expression "justification". We then end up either admitting two different notions of meaning (an unpalatable conclusion) or facing a patent circularity: our theory of meaning is grounded on our theory of justification – while our theory of justification is grounded on our theory of meaning.

---

1 Cf., for example, Dummett (1976), (1978) and (1979).
2 We will discuss neither the different varieties of semantic contextualism and epistemological contextualism, nor all their problems. For some of them, cf. Preyer and Peter (2005).
4 In what follows, we will use "assertion" instead of "belief". This terminological choice – uncommon in epistemology, where we normally speak of beliefs – allows us to grasp immediately some epistemological consequences of Dummett's notion of meaning. We will
Meaning as Justification and Foundationalism

Dummett interprets Wittgenstein's slogan "meaning is use" as follows: it is the justification of an assertion that constitutes the meaning of the assertion. In other words, the meaning of an assertion is given by the justification of the assertion, or by its conditions of assertibility. For our present purposes it is not necessary to consider the details of Dummett's thesis; it may, nevertheless, be useful to try to understand whether there is some specific theory of justification that he could adopt. In fact, if we do not succeed in finding this theory, our attempt to contextualize the notion of meaning through the contextualization of the notion of justification is doomed to failure – at least in a Dummettian perspective.

Foundationalism is the more traditional theory of justification. Its central idea is that assertions are divided into basic ones and derived ones. The former need no inferential justification, but have an immediate one. The latter are founded on the former and obtain their justification from them through deductive and inductive inferences. Basic assertions are useful to stop the regress of justification. Aristotle was the first to individuate the problem clearly.¹ According to him, regress:

1. can go on *ad infinitum*;
2. can stop with unjustified assertions;
3. can be circular;
4. can stop in immediately justified assertions.

Foundationalists choose (iv), and judge (i), (ii), and (iii) to be unsatisfactory.

When Wittgenstein tries to clarify his slogan "meaning is use", he says that "it is what is regarded as the *foundation* of an assertion that constitutes the sense of the assertion". However, in the official English translation we have him saying: "it is what is regarded as the *justification* of an assertion that constitutes the sense of the assertion".² This is of course the origin of Dummett's conception of meaning as justification. Wittgenstein's reference to "foundation" is important since he may in fact be interpreted as holding a foundationalist conception of justification – distinguishing between basic propositions and inferentially justified ones: "If you do know that *here is one hand*, we'll grant you all the rest. When one says that such and such a proposition can't be proved, of course that does not mean that it can't be derived from other propositions; any proposition can be derived from other ones. But they may be no more certain than it is itself".³ Furthermore, though

³Wittgenstein (1969b, 1).
he seems to leave more than one option open, Wittgenstein embraces the foundationalist answer to regress, i.e. (iv)\(^1\).

If now we try to understand whether there is some theory of justification that Dummett could adopt, we are inclined to believe that he has in mind foundationalism. On the one hand, as we have said, he derives inspiration for his theory of meaning as justification from Wittgenstein (and Wittgenstein adheres to a version of foundationalism). On the other hand, one of Dummett's strong convictions is that most assertions cannot be used as reports of immediate observations, and can be established only through an inference based on what has been observed; their justification must, therefore, be inferential.\(^2\)

It is thus utterly reasonable to claim that Dummett holds a foundationalist view about justification and distinguishes between basic assertions – justified by observation – and inferentially justified assertions.

**Contextualizing the Notion of Justification**

The problem of justification is present throughout the entire history of philosophy. We cannot but recognize that this history has been dominated by invariantism. According to this thesis there is one and only one epistemic standard, and therefore it is wrong to claim – for the same cognitive subject \(S\) and the same proposition \(p\) – that

(1) \(S\) is justified in asserting that \(p\) is true in one context, and false in another context. Conversely, the contextualist thesis admits the legitimacy of several epistemic standards that vary with context of use of (1): it is right to claim – for the same cognitive subject \(S\) and the same proposition \(p\) – that (1) is true in one context, and false in another context.

David Annis proposes a contextualist account of justification in a foundationalist framework. According to Annis, \(S\) must be able to meet certain objections couched in terms of precise epistemic aims – i.e. achieving true assertions and avoiding false assertions. Concerning a proposition \(p\), the epistemic claims of \(S\) may be objected to in two different ways: (a) \(S\) is not in a position to know that \(p\) is true; (b) \(p\) is false. Because we do not want to have conditions so strong that \(S\) cannot satisfy them, not every objection is possible or, at least, \(S\) is not required to answer every objection. Objections must be

---

\(^1\)Cf. Wittgenstein (1969b, 192). The problem is then to understand what he means by "immediately justified assertions". In fact, he writes: "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" (1969b, 204); and "I KNOW that this is my foot. I could not accept any experience as proof to the contrary.- That may be an exclamation; but what follows from it? At least that I shall act with a certainty that knows no doubt, in accordance with my [assertion]" (1969b, 360). Moreover, on at least one occasion, Wittgenstein seems to accept answer (ii): "At the foundation of well-founded [assertion] lies [assertion] that is not founded" (1969b, 253). These are, however, exegetical difficulties that do not concern us here.

\(^2\)Cf., for example, Dummett (2001, p. 136).
"based on the current evidence available", and "must be a manifestation of a real doubt where the doubt is occasioned by a real life situation". It may be said that S "is not required to respond to an objection if in general it would be assigned a low probability by the people questioning S". Obviously, these people must pursue the above epistemic aims – looking for truth and avoiding falsity – because, otherwise, their objections will not be appropriate, and S must reply in such a way as to produce a general agreement that the force of these objections entirely collapses or, at least, is considerably decreased. It is, however, obvious that S can also reply by noting that the objections are not appropriate, because they are not based on the actual available evidence, or because they are not the result of a real doubt, or because they do not pursue epistemic aims, or because they do not belong to the types (a) or (b).

The question "Is S justified in asserting that p is true?" is always relative to an issue-context or to a conversational context. For example, let us suppose that we are going to decide if Sophie – an ordinary person in an ordinary context – is justified in asserting that Crete monopolized the trade routes in Eastern Mediterranean during the so-called Minoan Age. We ask Sophie: "Why do you assert it?". We are satisfied if she answers that she has read it in an article about Crete in a newspaper and that newspapers are generally reliable, because we apply a rather relaxed epistemic standard – we are in an ordinary context. Of course, the same answer is not accepted if the context changes. In fact, let us suppose that Sophie is taking her bachelor examination in Eastern Mediterranean History. We do not judge her justified at all in her assertion if she appeals to her having read the newspaper, because in this new context we apply a rather elevated epistemic standard. So, with regard to an issue-context a subject can be justified in asserting a proposition p, and with regard to another issue-context the very same subject may not be justified at all in asserting the very same proposition. It is evident that the issue-context "determines the level of understanding and knowledge that S must exhibit, and it determines an appropriate objector-group". So, while in an ordinary context the appropriate objector-group is constituted by ordinary people and not by historians, in the above bachelor examination it is surely constituted by historians. Given a certain issue-context, if the appropriate objector-group asks S reasons for her assertion, this assertion is not a basic one in that context, because it will be derived from reasons and, therefore, from assertions that are meant to support it. In the above ordinary context Sophie’s assertion is obviously derived because the basic assertion is: "newspapers are generally reliable". But in a context where the reliability of newspapers is in question, the assertion "newspapers are generally reliable" will no longer be basic.

The regress problem seems solved, but there remains in fact a rather serious difficulty. Since S is not required to respond to an objection if it would be assigned a low probability by the appropriate objector-group, S is justified in asserting a proposition p only if:

---

3 Accounts of this difficulty are present elsewhere: cf., for example, Brady (1998).
1. the appropriate objector-group G does not ask reasons for S's assertion - because of the low probability assigned to an objection O: the assertion is contextually basic and justified;
2. G asks reasons for S's assertion - because of the high probability assigned to O: if S can adequately answer, her assertion is justified, though not contextually basic but obviously derived (S has to appeal to other assertions in order to reply to O).

It is important to notice that, regarding basic assertions, it is the appropriate objector-group G that assigns a low probability to O. Consequently, this appropriate objector-group G must assert that an objection O has a low probability. Let us call w the assertion "O has a low probability" made by G, and let us ask if G is justified in asserting w. According to the contextualism under consideration, we must take into account an appropriate objector-group G₁ in order to say that G is justified in asserting w. Again, we have two possibilities:

1. G₁ does not ask reasons for G's assertion - because of the low probability assigned to O₁: the assertion is contextually basic and justified;
2. G₁ asks reasons for G's assertion - because of the high probability assigned to O₁: if G can answer adequately, the assertion is justified, though not contextually basic, but obviously derived (G has to appeal to other assertions in order to reply to objection O₁).

But, of course, regarding basic assertions, also G₁ has to assign a low probability to O₁. Thus, also G₁ must assert that O₁ has a low probability. Let us call y the assertion "O₁ has a low probability" made by G₁ and let us ask if G₁ is justified in asserting y. We must once again take into account an appropriate objector-group G₂ in order to say that G₁ is justified in asserting y, and again:

1. G₂ does not ask reasons for G₁'s assertion - because of the low probability assigned to O₂: the assertion is contextually basic and justified;
2. G₂ asks reasons for G₁'s assertion - because of the high probability assigned to O₂: if G₁ can adequately answer, the assertion is justified, though not contextually basic, but obviously derived (G₁ has to appeal to other assertions in order to reply to objection O₂).

Our argument can obviously be reiterated with G₂ which requires the appropriate objector-group G₃, G₃ which requires the appropriate objector-group G₄, G₄ which requires the appropriate objector-group G₅, and so on, ad infinitum. And if this line of reasoning is reiterable, as it seems, the contextualism under consideration cannot solve the regress problem. In fact, while foundationalism wants the regress to stop in immediately justified
assertions – solution (iv) above – contextualism cannot but adopt a regress that
goess on ad infinitum – solution (i) above.

In order to overcome the above serious difficulty, we can perhaps adhere to
a view that does not mention the necessity of any objector-group, but appeals
solely to the features of the conversational context, as in Keith DeRose's view:

If you’re a foundationalist, then if you’re also a contextualist, you
may well come to think of the issue of which [assertions] are
properly basic (i.e., the issue of which [assertions] are justified to a
degree sufficient for knowledge independent of any support they
receive from other [assertions]), and/or the issue of how strongly
supported an [assertion] in the superstructure must be in order to
count as knowledge or as a justified [assertion], to be matters that
vary according to features of conversational context.¹

We must, however, specify in detail the features of the conversational
context; sooner or later, we may find "the appropriate objector-group" among
them. So, in order to block the regress and overcome the above difficulty, it is
better to turn to Michael Williams’ concept of methodological necessity. In
order to converse (or to carry out scientific research), we must restrict the
possibility of raising objections: it is not permissible to go on asking for
justifications, because this would induce us to ignore conversational
constraints:

In both science and ordinary life, constraints on justification are
many and various. Not merely that, they shift with context in ways
that are probably impossible to reduce to rule. In part, they will have
to do with the specific content of whatever claim is at issue. But they
will also be decisively influenced by the subject of inquiry to which
the claim in question belongs (history, physics, ornithology, etc.).
We can call these topical or, where some definite subject or
distinctive form of inquiry is involved, disciplinary constraints…
Disciplinary constraints fix ranges of admissible questions. But what
is and is not appropriate in the way of justification may also be
strongly influenced by what specific objection has been entered to a
given claim or [assertion]. So to disciplinary we must add dialectical
constraints: constraints reflecting the current state of a particular
argument or problem-situation.²

Contextualizing the Notion of Meaning

In section 3, we have sketched a plausible contextualist proposal regarding
justification. Our proposal has a crucial feature: it contextualizes

¹DeRose (1999, p. 190).
²Williams (1991, p. 117).
foundationalism, and foundationalism – as we have shown in section 2 – is the theory of justification Dummett is inclined to accept. At this point, the notion of meaning can be contextualized in a straightforward way. Meaning amounts to justification (as Dummett claims) and justification can be seen in a foundationalist perspective (as we have argued): if we can contextualize foundationalism in a reasonable way, then we can obviously contextualize the notion of meaning, thanks to the contextualization of the notion of justification. So, as promised in section 1, we obtain an original and innovative account of a contextualistic theory of meaning. However, we are well aware that we must face several major problems: let us consider some of them.

First of all, let us take into account the semantic value of assertions containing terms related to justification, as for example "to be justified". We have said that, according to contextualism, it is legitimate to state that

(1) S is justified in asserting that p is true in one context, and false in another context. Clearly, since the predicates "true" and "false" are indispensable for this kind of contextualism, we must appeal to the notion of meaning as truth-conditions, and not to the Dummettian notion of meaning as justification.

It is easier to realize this point if, taking a semantic perspective, we deal with assertions containing "is justified" as we deal with assertions containing indexicals. The interpretation of an assertion containing an indexical depends on the characteristics of the context in which the assertion is uttered. This means that the interpretation varies with the context of use:

(2) I am Greek is false if uttered by Alessandro (who is an Italian), and true if uttered by Alexandros (who is a Greek). According to Kaplan, we need to distinguish the "character" from the "content": in different contexts of use, the character of an assertion containing an indexical is constant, while the content changes; such a character is the linguistic meaning or, better, a function from contexts of utterance to contents, while the content is the sense or the proposition expressed by the assertion.\(^1\) What is, then, the character of (1)? According to DeRose, it is roughly the following: "S is in a good enough epistemic position with respect to p".\(^2\) And what is the content of (1)? It is how good an epistemic position S must be in to count as being justified in asserting that p - and this shifts from context to context. Let us now consider the following assertion:

(3) Sophie is justified in asserting that there is some excellent wine in this bottle. The character of "is justified" in (3) is invariable. The content varies with the context of attributor, and, in particular, with the epistemic position the attributor requires for the cognitive subject. If I am not a sommelier, but a normal person, I may truly state that (3) is true, if Sophie is in a certain epistemic position: for example, it is clear to her that the wine is of an "appellation controlée", that it is palatable and tasty. Of course such a position is not judged good enough if I am a sommelier, and so in this case I may state that (3) is false. The way in which the truth-conditions of (3) vary with context is not different at all from the way in which the truth-conditions of (2) vary

\(^2\)DeRose (1992, p. 922).
An Anthology of Philosophical Studies

with context. We cannot but admit that our way of contextualizing the notion of meaning through the contextualization of the notion of justification presupposes not only the notion of meaning as justification, but also the notion of meaning as truth-conditions — contra Dummett.

We have said that a contextualist account of meaning seems to presuppose a notion of meaning as truth-conditions. We can further argue in favor of this very point by noting that all the different theories of justification (foundationalism included) try to clarify precisely the meaning of the expression "justification", in spite of the well-know problems affecting the very notion of meaning. In this regard, what Alvin I. Goldman states is paradigmatic:

First, while there are doubtless severe theoretical problems concerning the notions of meaning and synonymy, there must be some substance to the commonsense notions suggested by these terms. Certainly we can distinguish better and worse definitions of a given word, whether dictionary definitions or definitions offered by casual speakers… So there must be some phenomenon of meaning that remains to be clearly elucidated. Second, although many philosophers preach the abandonment of analyticity, their practice sometimes belies their preaching. People do things very much like conceptual analysis even if they officially reject it. It is hard to do much in epistemology (or other branches of philosophy) without feeling constrained to do something like conceptual analysis.¹

Doing conceptual analysis amounts to presenting the necessary and sufficient conditions for key epistemological terms, the term "justification" included. The concept, i.e. the analysandum, is expressed by the pattern "S is justified in asserting that p", and the analysis by the pattern "S is justified in asserting that p if and only if…", where the dots must be replaced by the analysans, i.e. a list of necessary and sufficient conditions.

In that case, if we follow Dummett in adopting a notion of meaning as justification, and if, in epistemology, we want to define "S is justified in asserting that p", we cannot but end up with two notions of meaning: the notion of meaning as justification and the notion of meaning as truth-conditions. Since there is an obvious clash between the two notions, this result is not at all desirable. We face two possibilities.

A. If we choose to admit just one notion, and continue to preserve Dummett's notion, we cannot but opt for the notion of meaning as justification; since it is hard to do much in epistemology without doing something like conceptual analysis, we cannot but try to clarify the meaning of the term "justification". It is inevitable at this point – in the absence of the notion of meaning as truth-conditions – to run into circularity. We must admit that, in order to clarify the meaning of "justification", we have to resort to the very

¹Goldman (1986, p. 38).
notion of meaning as justification: therefore we presuppose the very notion we want to clarify the meaning of, i.e. the notion of justification.

B. Since the notion of meaning as truth-conditions allows us to understand the meaning of "S is justified in asserting that p" without running into any circularity, it seems that there is no other way than to resort once more to the notion of meaning as truth-conditions. Yet, since we cannot admit – for the reasons we have already considered – two different notions of meaning, i.e. the notion of meaning as truth-conditions and the notion of meaning of justification, it seems reasonable to give up the notion of meaning as justification. However, on Dummett's view, this rejection is impossible. As a matter of fact, he thinks that the notion of meaning as truth-conditions requires the notion of meaning of justification: when we make an assertion, we do not only imply that the assertion is true, but also that we have a justification for its truth.¹

If, by appealing to the notion of meaning as truth-conditions, we cannot avoid the appeal to the notion of meaning as justification, it is impossible to avoid the circularity underlined above. Obviously such circularity casts considerable doubts on the success of our attempt to contextualize the notion of meaning through the contextualization of the notion of justification.

A Brief Conclusion

In our paper, we have outlined an attractive way of contextualizing meaning, and we have pointed out its major shortcomings. While in Dummett's semantic theory the meaning of an assertion is given by its justification, in epistemology a theory of justification aims to clarify the meaning of the expression "justification". We ended up either admitting two different notions of meaning, or facing a patent circularity: our theory of meaning is grounded on our theory of justification – while our theory of justification is grounded on our theory of meaning. If we do wish to preserve our attempt to contextualize the notion of meaning through the contextualization of the notion of justification, and avoid the above circularity, we might then give up the practice of giving necessary and sufficient conditions for the expression "justification" and consider "being justified" as a natural kind or a natural property. Considering being justified as a natural kind or a natural property leads to an investigation of justification as a natural phenomenon, and consequently to a naturalization of the very notion of justification.² If we end up with a naturalized notion of justification, and at the same time maintain a notion of meaning as justification, obviously we must naturalize the notion of meaning as well. However, naturalizing the notion of meaning does not amount to contextualizing it.

²A strategy of this kind is clearly present in Kornblith (2002).
References


Part 3

Ethics and Political Philosophy
Can Political Liberalism Deliver Equality in the Social-Bases of Self-Respect?

Gerald Doppelt

Equality in the social bases of self-respect, as a requirement of justice, is reiterated in every one of Rawls’ major works.\(^1\) This notion is at the core of Rawls’ theory in that it accounts for (1) the content of his two principles of justice, how they distribute primary social goods, (2) how the principles are justified, (3) the reasons he gives for thinking they can be embodied in a well-ordered society, and (4) the most fundamental ‘political’ conception or ideal of persons implicit in our democratic traditions. What follows below are Rawls’ most important claims about self-respect:

1. Self-respect is a primary good, like income or liberty. It is an essential component in, and means to, every person’s ability to freely choose and realize some conception of the good life. Its distribution among individuals is a function of the basic structure of society. As such its distribution is a matter of political justice, not exclusively a matter of personal responsibility or justice between individuals.\(^2\)

2. Among the primary goods distributed by a just society, self-respect is “perhaps the main primary good”. Without it, one may reasonably believe that nothing is worth doing, rendering the other primary goods, and justice itself, useless. Thus, in determining the basic principles of a just society, we “would wish to avoid at almost any cost the social conditions that undermine self-respect”.

3. In modern society, the basic civil rights and political liberties of democratic citizenship constitute the social bases of self-respect. Thus, ‘equality’ in the social bases of self-respect demands ‘equality’ in basic civil and political rights (the first principle). It permits inequalities in socio-economic position provided that they work out to raise the material prospects of the worst-off, and do not

---


\(^2\)JF, p. 58-59.
undermine the fair worth of equal political liberties, or equality of opportunity (the second principle). Equality in the social bases of self-respect forbids the acceptance of a lesser liberty for all, or unequal liberty, for the sake of greater material benefits (the lexical priority of the first over the second principle). Thus, the primary good of self-respect plays a key role in justifying how all the other primary goods ought to be distributed.\(^1\)

4. The fact that Rawls’ principles deliver ‘equality in the social bases of self-respect’ provides a major justification for taking them to satisfy the ideal of a well-ordered society. Principles of justice are evaluated by treating them as the publicly recognized standards of a society which ideally fulfills them and then making an empirical determination of how stable and well-ordered the result will be.\(^2\) Because Rawls’ principles provide equality in the social bases of self-respect, while utilitarianism is far more precarious in this respect, the former is better justified than the latter, on the key criterion of well-orderedness.\(^3\)

5. Justice as fairness depends on claims about the social bases of self-respect and which conception of justice is in fact the one which yields the most well-ordered society. These are not methodological weaknesses. Because, the justification of any conception of justice should depend on “the conditions of life as we know it” and “a certain theory of social institutions”.\(^4\)

**Problems with Rawls’ Paradigm of Self-respect**

This account hopefully makes it clear just how provocative and important self-respect is in Rawls’ theory of justice. There is little reason to doubt that people gain self-respect when their equal rights of democratic citizenship are respected, or, that their self-respect is compromised or undermined when these rights are denied. Nonetheless, given “the conditions of life as know it”, there seem to be other powerful social bases of self-respect, and inequalities of respect, ignored by Rawls’ paradigm of democratic citizenship. Drawing on the voices of various critics, there seem to be social injuries to persons’ self-respect, rooted in inequalities of class, gender, and race, that are not accommodated by Rawls’ paradigm of self-respect, or the principles of justice that embody it. But do these sorts of criticisms stretch the notion of ‘social’ bases of self-respect beyond the pale of justice? Do they underestimate the impact of Rawlsian justice on inequalities of class, gender or race? In this essay, I focus on problems raised by class.

---

4. TJ, p. 58-60, 454.
How does Rawls regard problematic inequalities of class? His second principle of justice distributes several primary goods associated with class: income, wealth, the powers and prerogatives of authority, and presumably, occupational status. As it is usually understood, the principle justifies inequalities in such goods provided they work out to raise the material prospects (income, wealth) of all, especially the worst-off. The most economically efficient organization of the economy may involve substantial inequalities in the powers and prerogatives of authority, occupational status, and even income. Indeed, it may involve unemployment or underemployment for some and unpaid homemaking labor for others. For Rawls, all of these differentials can be just, provided that there are welfare-state redistributive mechanisms in place to assure that the economic dividend of this efficiency is used to raise the income of the worst-off, higher than it would otherwise be.

We can reasonably believe that this distributional justice would eradicate the indignity, humiliation, and degradation of poverty, homelessness, lack of health care, etc. in most democracies today; because with a decent standard of living for all, such phenomenon presumably disappear. But there may still be large disparities between persons based on whether they are employed, the conditions of their work, their relative occupational status, the power they have or lack over their own work or that of others, etc. We know that independently of income, these inequalities enter into the social bases of self-respect and generate low self-esteem among the unemployed, the unskilled, those whose work is mindless, monotonous, self-stultifying, rigorously controlled or monitored by supervisors, etc. Rawls’ paradigm of self-respect as citizenship ignores these social dynamics and thus may reproduce and justify them. Indeed, the fact that some of this work is low-wage compounds the injury to one’s sense of personal worth. In Rawls’ system, this dynamic may remain, because there will still be the distinction between those whose incomes embody what they individually earn on the open market, and those whose incomes derive from the welfare-transfers of political redistribution. Of course, whatever inequalities of income and wealth are justified, on Rawls’ principles, may also generate inequalities in the social bases of self-respect. In a highly materialist, consumerist, and competitive culture like ours, disparities in standard of living per se can imply a sense of personal failure. Do we have any

\[2\] TJ, p. 76-79, 97, PL, p. 5-7, 291, JF, p. 42-43. Note that in PL and JF, Rawls formulates the 2nd, or ‘difference principle’ to justify inequalities that ‘are to the greatest benefit of the least advantaged members of society’, PL, 42-43. As I indicate elsewhere in this essay, such inequalities can only be justified if they are compatible with fair equality of opportunity and the fair worth of the equal political liberties.
\[3\] TJ, p. 78, 87, 258-284.
reason to believe that all these inequalities will function differently in Rawlsian society

First Response: Tinkering with the Principles

At the end of *A Theory of Justice*, Rawls worries that ‘To some extent men’s sense of their own worth may hinge upon their institutional position and income share’.¹ He characterizes this as ‘an unwelcome complication’. In response, he suggests that in theory, we can redefine ‘the primary good of self-esteem’ as an independent variable and include it as one component in ‘the index’ for measuring worst-off and better-off.² In determining which socio-economic inequalities are just, we would then have to take into account its’ effects on self-respect, as well as on income, the powers of prerogatives of authority, and occupational status; and somehow weigh each against the others to reach an ‘on-balance’ judgment.

It is easy to see why this is an ‘unwelcome complication’. Once self-respect is demoted to but one more component of economic advantage, then gains or losses in self-respect can be compensated by increments in income or wealth. Such a move abandons Rawls’ core commitment to equality in the social bases of self-respect and the whole founding intuition that self-respect is ‘perhaps the main primary good’. But this core commitment plays key roles in justifying his theory’s requirement of equal liberties, the lexical priority of equal liberty over economic benefit, and the argument that his theory satisfies the test of well-orderedness better than utilitarianism.

Of course, there may be better ways of tinkering with Rawls’ principles. Rawls limits permissible socio-economic inequalities to those which do not undermine the fair worth of equal political liberties; i.e. the ability of citizens to use their political rights to effectively participate as equals in the democratic process.³ Given Rawls’ paradigm of self-respect as equal rights, this move is tantamount to limiting permissible socio-economic inequalities to those which do not undermine equality in the social bases of self-respect. But we could generalize this move to better accommodate the above ‘unwelcome complication’. If certain kinds of disparity in occupational status, the powers and prerogatives of authority, income and wealth, and source of income (paid work vs. welfare transfer programs) generate inequalities in the social bases of self-respect, they can be ruled out as unjust. We could reframe the second principle so that equality in the social bases of self-respect enjoys a lexical priority over the other socio-economic goods in determining better-off and worse-off. While this reformulation would be more consistent with Rawls’ basic argument, it may well destabilize his working model of a well-ordered society as a welfare state democracy with private property in the means of production, and the appropriate policy of income redistribution.

¹ *TJ*, p. 546.
² *TJ*, p. 546.
In any case, the turn in Rawls thought to ‘political liberalism’ provides a more promising response to these problems I have raised for his paradigm of self-respect as democratic citizenship.

Second Response: The Social Bases of Self-Respect as a Normative Notion

So far, I have read and criticized Rawls’ account of the social bases of self-respect as an empirical notion based on ‘the conditions of life as we know it’ or, as he also puts it, on ‘a certain theory of social institutions’. This is a natural reading of much of what Rawls says and a natural basis of criticism.

But, we may gain a deeper philosophical defense of Rawls’ argument if we treat his account as a normative one concerning equality in the proper, or most reasonable, social bases of self-respect, in an ideally just society. Given Rawls’ ideal of well-orderedness, there would still be empirical constraints on this normative notion concerning its probability of realization. But even so, equality in the **proper** social bases of self-respect – those which ought to (and would) shape self-respect in an ideally just society – does not require equality in the actual social bases of self-respect that prevail in any existing liberal-democratic society such as the USA. This is a benefit, because some prevailing social bases of recognition and self-respect may depend on unreasonable, even oppressive standards. To take an obvious example, girls and women may gain or lose recognition and self-respect on the social basis of how well their faces, bodies, and behavior conform to powerful social stereotypes of the ‘feminine’ promulgated by pervasive institutions of media. We can characterize such standards as unreasonable or oppressive if they distort the proper social bases of recognition and self-respect, on the normative account proposed by Rawls. Rawls’ political liberalism in his later work explicitly develops a normative conception or ideal of persons as free and equal, possessing two highest order moral powers: (1) the capacity to freely develop and realize a self-chosen conception of the good life, and (2) the capacity to recognize and respect the claims of others, based on their equal possession of the same capacity described in (1).\(^1\) With this ideal in place, we can derive a normative account of the proper social bases of self-respect as the equal rights and liberties of democratic citizenship. For in a just society these rights and liberties serve to protect persons’ freedom and equality and enable citizens to express respect for one another as free and equal persons, with the requisite moral powers.\(^2\)

We are thus lead from our concerns with Rawls’ paradigm of self-respect to the ‘democratic’ idea(1) of persons as free and equal that takes center stage in his turn to ‘political liberalism’. Rawls argues that this idea of persons is not to be understood as a metaphysical doctrine concerning the nature of personhood nor as a comprehensive religious, ethical, or philosophical view concerning the

---

\(^1\) PL, p. 18-19, 29-35.

\(^2\) PL, p. 318-320, JF, p. 4-5, 18-24, 199-200. This ideal of persons is presented in TJ (p. 179-180, 252, 505) as a Kantian conception, which Rawls rejects later as too metaphysical or comprehensive a conception to meet the needs of a “political” liberalism.
highest purposes and ends that guide all aspects of persons’ lives as a whole. Rather, it must be constructed as an exclusively political idea(l): one which is entirely compatible with the plurality of different comprehensive views of life that deserve respect in a liberal society, and can flourish reasonably, within the bounds of political justice.\(^1\) The democratic idea(l) of persons as free and equal can be constructed as a suitable ‘political’ conception for the following reasons. It can be justified independently of any particular comprehensive view, because it expresses widely shared values implicit in longstanding democratic traditions and institutions; fundamental ideas common to otherwise divergent comprehensive views and the people who hold them. It can be applied in a rigorously circumscribed way to the ‘basic structure’ of constitutional and legal institutions, and limited to persons’ relations as citizens. As such, the democratic ideal will justify the equal rights and liberties of citizenship as the proper social bases of self-respect in an ideally just, well-ordered society. People in that society will internalize its democratic conception of persons providing equality in the social bases of self-respect.

Nonetheless, there is still good reason to worry that the democratic idea(l) of mutual respect among person as citizens is too abstract and idealized to get what Rawlsian justice requires: the sense that what we do in everyday life is worthwhile, the concrete self-esteem and self-confidence that his vision of a just, well-ordered society promises. Surely, our pervasive social relations as economic actors, and as gendered members of families, imply their own normative idea(l)s of persons and notions of the proper social bases of self-respect. Some of these normative notions conflict with the democratic idea(l) and make equality in the social bases of self-respect seem either impossible, or undesirable, at least as these normative notions are now implicitly understood. But can they be excluded as ‘non-political’, therefore inappropriate subjects of concerns for a ‘political’ liberalism; or subordinated to the democratic ideal which is more ‘fundamental’ in our ‘public political culture’?

Underlying political liberalism lies Rawls’ insight that “society’s main institutions and their accepted forms of interpretation, are seen as a fund of implicitly shared ideas and principles.”\(^2\) Rawls obviously regards the economy and family as among ‘society’s main institutions’ and part of ‘the basic structure’ regulated by political justice.\(^3\)

But they are not seen by him to have their own traditional or accepted ‘forms of interpretation’ which imply a fund of implicitly shared ideas and principles in their own right. On the contrary, if these institutions constitute historical traditions based on implicitly shared idea(l)s just as much as the state (constitutional law, etc.), then they are central parts of the public culture and cannot be excluded from the sphere of the political. So I will argue.

In fact, our economic institutions rest on a Lockian idea(l) of persons as ‘free and equal’ participants in competitive market relations who act as producers and consumers of scarce values and, in so doing, affirm their worth

\(^1\) PL, p. 11-15, 22-35.
\(^2\) PL, p. 14.
\(^3\) PL, p. 258.
as persons. This conception foregrounds the ‘moral powers’ of persons that enable them to effectively enter into orderly, civil, market exchange and freely and fairly determine what they will earn, achieve, deserve, and be. Property rights, including property in one’s own person, constitute central social bases of recognition and self-respect on the Lockian idea. Clearly there are more and less democratic expressions of Lockian individualism, depending on historical interpretations that are more exclusionary (a zero-sum game where ‘winners’ require ‘losers’) or more inclusive (full employment, meaningful labor, wider professionalization of occupations). Thus, the Lockian ideal will justify at least some of the above empirical connections between self-respect and persons’ positions in the division of income, labor, power, occupational status, professional activity and paid employment. They will count as proper and reasonable social bases of self-respect. Political liberalism will not want to ignore these ‘unwelcome complications’, because the Lockian idea(l) of persons cannot be dismissed as non-political. Rather, any ‘political’ theory of justice will need to reconcile the democratic idea(l) of free and equal persons with the Lockian idea(l). Whether we evaluate Rawls’ paradigm of self-respect as a purely empirical or normatively ideal conception, political liberalism will need to treat the challenge posed by the conflicting conceptions of persons built into our most basic institutions and traditions. For they yield a serious measure of normative disarray, as well as inequality, in the social bases of self-respect.

Yet, if we hold on to Rawls’ claim that justice requires equality in the social bases of self-respect, and expand his political liberalism to encompass the deep politics of all basic institutions (state, economy, family, etc.), then much of his framework may yet guide the way to political justice. The fundamental idea is that political justice is based on ideals built into our shared public culture and traditions. Were we not swayed by our own particular comprehensive view or self-interest, what principles of justice would we choose to place everyone in the best position to realize these ideals and create a well-ordered society on their bases? The thought that equality in the social bases of self-respect is at the heart of this quest for justice, is not one we would do well to give up.

**Gender and Inequalities in the Social Bases of Self-Respect**

While I have so far focused on economic institutions and self-respect, the family raises issues of its own. Feminist philosophies have argued that gender forms an institution of social domination and subordination, which pervades the relations of males and females in family life, the workplace, institutions of government, and public culture generally. Here, I can only pick up one strand of this argument that bears directly on my treatment of political liberalism.

Susan Okin seeks to develop a Rawlsian feminism, which tests the limits of political liberalism, as Rawls defines them. She argues that properly understood, Rawls’ principles of justice imply a restructuring of the gendered division of labor, power, opportunity, economic independence-or-vulnerability,
and self-respect, built into the traditional institution of the family. Reformulated in my terms, Okin identifies a contradiction between Rawls’ political ideal of persons as free and equal, and the gendered conception of persons that prevails in the institution of the family and the ways children are socialized by it. There are several dimensions to Okin’s argument, aimed at simultaneously holding onto Rawlsian justice as fairness and principles of justice, while drawing out implications that bring marriage and family into the basic structure in a more radical way than Rawls does. One dimension makes the claim that as a matter of empirical fact, Rawls’ political ideal of free and equal citizens, equal civil and political liberty, equality of opportunity in economic life, and a high social minimum for all, cannot be realized in society as a whole, without a restructuring of the gender relations of family life; gender inequities in the family will (do) simply translate into inequalities in opportunity, the fair worth of political liberties, and the social bases of self-respect in the larger society. This dimension of Okin folds neatly into Rawls’ theory of justice, while altering its institutional implications. A second more radical dimension makes the claim that persons are not free and equal unless they are free and equal in the relations of family life; e.g. Equal in decision-making prerogatives, autonomy, the duties of childrearing, the prerogatives of breadwinning, or at least free from having these determined by gender-roles, as against individual proclivities. This dimension of Okin alters the very content of Rawls’ primary goods and his principles of justice, not just their institutional application. For, equal rights and liberties, the power and prerogatives of authority, and equality in the social bases of self-respect now attach directly to the institutional positions of women and men in the family. The substance of justice alters, straining its characterization as a ‘Rawlsian feminism’. In later works, Rawls’ response to Okin’s argument is unequivocal. As long as no person’s basic civil and political rights are violated, marriages and families may be structured in very different ways without violating the ‘political’ idea(l) of persons, as free and equal. He treats marriage as less a powerful traditional institution in its own right, and more akin to any other voluntary, private civic association or comprehensive view in civil society. Political justice constrains any particular family or marriage in the same way as any particular comprehensive view or way of life is constrained; it is not reasonable if it violates the equal rights of persons as citizens, as formulated by Rawls’ two principles of justice. By implication, families may be structured on the basis of very different religious or philosophical views concerning sex-gender roles, and the proper relations of men, women, and children, all within the limits of the ‘political’ conception of justice. As in the case of our economic relations, our relations as members of families are not seen by Rawls as shaped by institutional tradition, public culture, and a fund of implicit ‘political’ idea(l)js of persons. I would challenge this exclusion of the family as a non-political institution.

---

2JF, p. 162-168.
While the family undergoes change, just as the democratic state and law does, there can be little doubt that it is a basic social institution governed not simply by law but by social mores, cultural conventions, and tradition-bound rules. Barbara Ehrenreich characterizes the dominant institution in mid-century America as the family-wage system, marked by a division of labor, power, and duty involving male breadwinners and female homemakers; as well as ideals of persons that tie manhood to the ability and willingness of every man to be a successful breadwinner for wife and children, and womanhood, to the ability and willingness of every woman to be an effective homemaker for husband and children. Ehrenreich shows that this was not only the way the institution worked, but also the normative conception articulated by many aspects of our public culture. Three decades of challenges to the family-wage system have altered the shape of families and brought ideals of manhood and womanhood into the center of political debate. While traditional gender ideals have more to contend with, from women and men bringing democratic and Lockian aspirations to gender relations, arguably they still are implicit in family relations, in attenuated form. In any case, the positions people occupy in family relations, in so far as they are defined by the institution, constitute social bases of self-respect, widely regarded as proper and reasonable.

What are the implications of expanding political liberalism to recognize that as an institution, the family does rest on a political idea(l) of persons. Most obviously, equality in the social bases of self-respect becomes problematic if a powerful component of our public culture (traditional institutions of marriage and family) justifies, as proper, different social bases of self-respect for women and men (husbands, wives, fathers, mothers, sisters and brothers, grandmothers and grandfathers), to the detriment of women. Once we become aware of this, isn’t it reasonable to simply reject the gendered conception(s) of persons implied in traditional family institutions, in favor of the democratic conception of persons as free and equal citizens?

The ideal(s) of persons implicit in the tradition of the family ties persons to the roles they play, and the powers or virtues they exhibit, in the work of families- care-giving, breadwinning, homemaking, decision-making, child-rearing, life-planning, etc. These activities and roles can be democratized so that they are each equally proper as the work (duty, or pleasure) of men and women and equally, social bases of self-respect for both. Once political liberalism rethinks the conflicting ideals of persons implicit in major institutions, we can seek to reconcile them, with the guiding aim of equality in the proper social bases of self-respect.

Understood in this way, political liberalism will not think of the social bases of self-respect as exclusively primary goods or the institutional positions that provide them. Political ideals or standards of personhood lie beneath social bases of recognition and self-respect, creating both the tensions and possibilities for the political theory of justice we inherit from Rawls. Seen from this perspective, the task of political liberalism is not simply to recover the ideals of personhood implicit in basic institutions, but also to revise the

---

standards of social recognition which generate unreasonable scarcities in the social bases of self-respect. We need to extend the Rawlsian notion of ‘public reason’ to encompass the cross currents of conflicting cultural paradigms of respect for persons.\footnote{Rawls, John. “The Idea of Public Reason Revisited”, \textit{The Law of Peoples}. (Cambridge and London, Harvard University Press. 1992), p. 129-181.} Political justice is not just a matter of a more just distribution of well-recognized social goods. It requires that we rethink what is involved in being a full person, in a community of free and equal persons recognized as such in the everyday life of a recognizably just society.
All people have human rights and, intuitively, there is a close connection between human rights, needs, and autonomy. At least this paper will assume that human rights protect individuals’ ability to meet their basic needs and live autonomous lives. Accounting for the connection between human rights, needs, and autonomy is more difficult than one might expect, especially on the traditional kinds of human rights theories (even those endorsing ‘positive’ social and economic rights) (Wenar, 2005; Fagan, 2006). On many traditional theories, human rights protect (at least a subset of) individuals’ important interests. Such theories are well suited to account for the fact that human rights protect individuals from dire need. Even the non-autonomous have some needs, which constitute some of their important interests. But because these theories sometimes say autonomy is not constitutive of the interests human rights protect they can fail to capture the close relationship between human rights and autonomy. Interest based theories can also fail to protect autonomy for other reasons (e.g. see (Raz, 1998; Hassoun 2008a)). Other traditional human rights theories protect individuals’ autonomy. These theories avoid the problem sketched above. But, purely autonomy-based theories usually cannot explain the universality of rights since some people lack autonomy. Some are not even potentially autonomous. The non-autonomous including the very young and severely disabled lack human rights on these theories (e.g. see (Griffin, 2006)).

1 The author would like to thank Gillian Brock, David Braybrooke, Thomas Pogge, Thomas Christiano, Michael Gill, Jerry Gaus, Keith Lehrer, Susan Wolf, Allen Buchanan, Dale Dorsey and audiences at the University of Arizona, ATINER 3rd International Conference on Philosophy Northwest Philosophy Conference, and Kwazulu-Natal University for helpful comments, suggestions, and/or discussion. She apologizes to anyone she has so carelessly forgotten to mention. Finally, she is terribly grateful for the support she received from the American Association of University Women, the University of North Carolina Chapel Hill, Duke University, and the Earhart Foundation during the course of the project.

2 It is beyond the scope of this paper to argue for this conclusion. For compelling arguments that human rights must protect individuals’ ability to meet their needs see: (Buchanan, 2004). For compelling arguments that human rights must protect individuals’ ability to secure autonomy see: (Griffin, 2006).

3 Griffin actually grounds his account in a more robust conception of autonomy than the one this paper will use (what this paper will call personhood). Furthermore, he might say that he is grounding rights in individuals’ interests since he supposes that people have an urgent interest in personhood. Nevertheless, his theory has the problem suggested here because he tries to ground all human rights only in this interest.
Furthermore, if human rights only protect individuals’ autonomy, human rights can be fulfilled and yet some can be left in dire need. So, some of the best known attempts to justify human rights either cannot appropriately connect human rights and autonomy or cannot account for the human right of all to meet their basic needs (Fagan, 2006; Hassoun, 2008a; Raz, 1998; Griffin, 2006). Any successful human rights theory must avoid this dilemma. It is possible to avoid this dilemma if (1) human rights, whatever else they do for people, protect individuals’ ability to live minimally good lives, (2) autonomy is necessary for and partly constitutive of such a life, and (3) people need all and only what will enable them to live such lives. This paper makes a preliminary case for (1) by showing that (2) and (3) hold. For, if (2) and (3) hold, a theory of human rights on which human rights protect individuals’ ability to live minimally good lives avoids the dilemma sketched above. If this argument goes through, then there is reason to take seriously a theory of human rights on which human rights protect individuals’ ability to live minimally good lives. Traditional human rights theories must embrace something like (1)-(3) if they are to prove adequate or a new theory of human rights may be necessary. In any case, the next section sketches an account of autonomy. The third section defends (2). The third and fourth sections together defend (3). The final section concludes.

Autonomy

The idea that people must be able to freely shape their lives is central to most accounts of autonomy (Nussbaum, 2000, 72). Though more may be necessary for even the most minimal sort of autonomy as a pre-requisite for free action, one must at least be able to reason about, make, and carry out simple plans on the basis of one’s desires. Let us say a bit about each of the conditions for this quite minimal kind of autonomy now.

First, to secure autonomy one needs some instrumental reasoning ability. Some hold demanding conceptions of autonomy on which reasoning requires much more. Kant, for instance, believes reason demands that one acknowledge the categorical imperative as unconditionally required (Hill, 1989; O’Neill, 1986). Many deny the existence of such an imperative. Most people can agree,

---

1Technically something might be partly constitutive of something which is partly constitutive of a minimally good life without being partly constitutive of a minimally good life. But we will argue that the requisite part of personhood is part of a minimally good life and autonomy is (in its entirety) a part of personhood.

2The following (implicit premises) should be uncontroversial, if not analytic: 1a. If autonomy is necessary for and partly constitutive of a minimally good life and human rights protect each individual’s ability to live a minimally good life, then human rights protect each individual’s ability to secure autonomy. 2a: If people need all and only what will enable them to live a minimally good life and human rights protect each individual’s ability to live a minimally good life, then human rights protect each individual’s ability to meet their needs.

3Those concerned to defend interest-based theories might have to argue, for instance, that people have an interest in x if x is partly constitutive of a minimally good life.
however, that autonomy requires instrumental reasoning ability. Such reasoning is essential for making simple plans and shaping one’s life.

To be able to make simple plans one must have some internal freedom, even if externally constrained. One must be able to adjudicate between one’s desires. To make sense of this idea, one might give a decision-theoretic analysis of planning in terms of creating a consistent preference ordering. Alternately, one might cash out planning ability in terms of ordering one’s ends (Rawls, 1971). Because these are standard moves in the literature on autonomy, however, we need not consider the ability to make simple plans further here.¹

Finally, to carry out simple plans one must be able to act on one’s consistent desires or well-ordered ends. One needs some external and internal freedom. One must not be completely prevented from acting. One must also have some internal control over one’s body.

Consider how autonomy is impaired when one cannot reason about, make, and carry out simple plans on the basis of one's desires. Suppose that Aida has a head injury that sometimes makes her delusional. Suppose that when she is delusional she cannot reason or plan. Aida might ask to speak to her husband one moment and then refuse to see him the next. Aida's desires are like trains on a busy track going this way and that, unconstrained by traffic signals.² Aida is not autonomous.

Contrast Aida’s case with Emal’s. Suppose Emal is a devout Muslim. He wants to live his whole life according to his faith. Occasionally he wants to drink with the other young men who live in his neighborhood. Suppose Emal can reason about, formulate, and carry out simple plans and so carries out a plan to go to the mosque and pray whenever he wants to drink with the other young men in his neighborhood. Since Emal can reason about, make, and carry out such plans on the basis of his desires, he is autonomous.³

The kind of autonomy explicated in this section is minimal. So, it should be able to secure broad assent. If one can make the case that human rights should protect a more robust conception of autonomy, the account that follows can be expanded.⁴ But in the absence of such arguments, let us adopt this conception of autonomy. The next section suggests that people need all and only what will enable them to live minimally good lives. The fifth section argues that the kind of autonomy sketched above is necessary for and partly constitutive of such a life. Finally, the concluding section considers how adopting this account of needs provides a reason to accept a theory of human rights on which human rights protect individuals’ ability to live minimally good lives. For, then, it is possible to avoid the dilemma with which we started -- everyone will have a human right to the protection of his or her autonomy and needs.

¹For instance, see (Bratman, 2005).
²This analogy is from (Feinberg, 1973).
³The conditions for autonomy I have set out do not prevent one from acting from poor reasons (e.g., wishful thinking). If one thinks this is not compatible with autonomy, additional criteria for autonomy will be necessary to rule out this possibility.
⁴The paper will only face a problem if one can both argue that human rights should protect a more robust conception of autonomy and that such autonomy is not necessary for a minimally good life.
An Anthology of Philosophical Studies

Autonomy and Needs

A few theorists have suggested something like the minimally good life account of needs on which people need whatever will enable them to live minimally good lives (Anscombe, 1958). None has cashed out the account. We might begin this project by appealing to the account of autonomy above. We will argue that autonomy is necessary for and partly constitutive of a minimally good life (Nussbaum, 2000, 72). This will also illustrate a general argumentative strategy by which one could show that other things are part of a minimally good life -- further cashing out this account of need.

Consider, first, why autonomy (understood here as just requiring reasoning and planning ability) is necessary for a minimally good life. Rewarding struggle, deep understanding, good relationships, significant achievement, virtue and so forth are some of the things that make a life go minimally well. Each of these things requires autonomy. People must be able to reason about, make, and carry out simple plans on the basis of their commitments to create and maintain good relationships. People must, for instance, be able to reason about, make, and carry out plans to talk with their friends and families for their relationships to flourish. Reasoning and planning are also necessary for developing important skills and character traits, deep understanding, significant achievement, and so forth. So, autonomy is necessary for a minimally good life.

Autonomy is also partly constitutive of such a life. This is because personhood is, partly constitutive of a minimally good life and autonomy is partly constitutive of personhood. Consider, first, why personhood is partly constitutive of a minimally good life. As persons we ‘have a conception of ourselves and of our past and future. We reflect and assess. We form pictures of what a good life would be, often, it is true, only on a small scale, but occasionally also on a large scale. And we try to realize these pictures’ (Griffin, 2006, Ch. 2). These conditions for personhood are also conditions for a minimally good life. To live a minimally good life one must be able to hope and dream, to pursue one’s goals and carry out projects, to live life on one’s own terms. Those who lack a conception of being a self, persisting through time, with a past and future cannot hope or dream. Those who never pursue their conception of a good life cannot achieve their goals or carry out projects. Hence personhood is partly constitutive of a minimally good life. Consider next why autonomy is partly constitutive of personhood. Recall that autonomy requires the ability to reason, make, and carry out simple plans on the basis of one’s desires. These conditions for autonomy are also conditions for personhood. To reflect and assess in the way that personhood requires one

1Some have suggested autonomy-based accounts of need. On such accounts, people need whatever will enable them to live autonomous lives. If this is correct then, on an autonomy-based theory, human rights will secure for people everything they need. The problem with this account of needs, however, is similar to (and does nothing to help resolve) the other problem with many autonomy-based theories of human rights. Some people cannot secure autonomy, but even these people have some needs (and rights).
must be able to reason. To pursue one’s conception of a good life, as persons do, one must be able to make and carry out simple plans.

We can also see that autonomy is necessary for and partly constitutive of a minimally good life via examples. Suppose Aefa drifts through life making one choice then another randomly or letting others choose for her. Suppose that Aefa has not freely chosen to drift. He simply cannot reason about, make, or carry out plans. Aefa cannot shape his own life. He does not choose consistently enough to attain most of the things he desires. He may end up subject to another’s will. Even if, by chance, Aefa secures many valuable things, his life will still lack an important kind of value. His life will be like a prize won accidentally (Raz, 1998). Aefa cannot live a minimally good life because his life is not truly her own.

One might object that some can have minimally good human lives without autonomy. The very young and severely disabled lack autonomy but may have such lives. They may experience joy and sadness, pleasure and pain, music and light (Kittay, 2005). Autonomy is not necessary for or even partly constitutive of such a life.

Disability theorists have convincingly argued that there may be a lot of value in a life without autonomy (Kittay, 2005). Non-autonomous people, like the rest of us, deserve respect and care. But the fact that it is valuable to experience emotions, music, and light provides little support for the claim that those who cannot reason and plan can live minimally good lives. Rather, it is such a great loss if the extremely disabled cannot secure autonomy precisely because everyone needs autonomy to live a minimally good life (though people also need other things to live such lives).

A different objection is that monks can live minimally good lives although they freely abdicate their autonomy. One might argue, for instance, that monks lack autonomy since they must follow the rules of the monastery and give up all of their worldly possessions. But, one might maintain, many monk’s lives are minimally good.

This objection misunderstands the kind of autonomy that we have argued is necessary for and partly constitutive of a minimally good life. One need only be able to reason about, make, and carry out some simple plans to have this kind of autonomy. So, most monks have this kind of autonomy. Most monks carry out all kinds of simple plans – to read, write, eat, pursue their ideal of a religious life, and so forth.

So far, this paper has argued that autonomy is necessary for and partly constitutive of a minimally good life. So it should be clear how one might argue that something else is necessary for or partly constitutive of a minimally good life. By arguing that autonomy is necessary for and partly constitutive of a minimally good life, however, this paper has also added some content to the minimally good life account of need. We can say, for instance, that people need

---

1 If, however, disability theorists insist on reserving the term “minimally good human life” for the valuable lives the extremely disabled (and very young) can live, then they can simply call the sort of life at issue in this account the “minimally good and autonomous life.” The issue may just be semantic.
An Anthology of Philosophical Studies

whatever will enable them to secure autonomy on this account. But why think people need all and only what will enable them to live minimally good lives? The next section motivates the minimally good life account of needs by considering the alternatives. The final section shows how this account provides the key to avoiding the dilemma facing many traditional human rights theories.

Justifying the Account

Some people need only a little food and water. Others (e.g. pregnant women) need much more. Some do not need expensive medicines or health care. Others (e.g. AIDS victims) require a lot of medical aid. Any good account of needs must accommodate differences in individuals’ needs resulting from differences in individual constitution and from the fact that individuals occupy different positions in society (Brock, 1998; Frankfurt, 1988). That is, a good account should capture all of the things each person needs without including anything someone does not need. It is plausible that the minimally good life account of needs fulfills this desideratum. This section will argue, however, that the two main competitors to this account -- Harry Frankfurt’s harm and David Braybrooke’s social role accounts -- do not fare so well.

Before beginning, however, it is important to note that the accounts we will consider are not designed to fulfill the above desideratum for a satisfactory conception of needs. Harry Frankfurt really only intends to account for the presumptive force of needs (Frankfurt, 1988). David Braybrooke wants to give an account of what people need that can inform public policy (Braybrooke, 1987). Still, Frankfurt and Braybrooke have the best developed accounts of needs so it is worth seeing if their accounts can fulfill the above desideratum. If they cannot, and we are right that the minimally good human life account of needs fares better, the minimally good human life account should be, at least provisionally, accepted.

Harm-Based Accounts

Harry Frankfurt in The Importance of What We Care About argues that we need the inescapably necessary conditions for avoiding harm. He says people need those things that are ‘necessarily necessary for avoiding harm’ (Frankfurt, 1988, 112). Frankfurt does not give a complete account of harm, but he says one is harmed if one is made worse off than before. He also claims that one’s

1One might argue for this desideratum by suggesting that a good account must provide a plausible basis for what a decent society must enable its members to secure. Unfortunately, it is not clear that would be the most satisfactory defense here. We are, after all, trying to argue for an account of human rights by showing that it will protect each individual’s ability to secure what they need. Still, the desideratum is independently plausible, and can probably be defended in another way. Finally, even if it cannot be given an independent defense, this paper may still illuminate important connections between rights and needs (as well as autonomy and a minimally good life).
situation must improve for one to avoid harm, if the only way to keep one’s situation from becoming worse is to make it better. Finally, Frankfurt says one is harmed if one remains in a bad condition because more of a bad thing is worse than less of it (Frankfurt, 1988, 110).

There are several problems with Frankfurt’s account. First, Frankfurt’s notion of harm is too inclusive. Intuitively, one may not be harmed if one’s bad state merely persists. One may not be made worse off than before by remaining in a bad state that does not become worse. Suppose, for instance, that Zahra is in the early stages of Parkinson’s syndrome, a degenerative disease. Her memory is affected but she can still feed herself. Suppose, further, that she is given a new medication that stabilizes her condition. Even though her bad condition remains the same (her memory is still poor), it seems that the medication has been helped, not harmed, Zahra.

Perhaps Frankfurt could say that when Zahra’s degenerative condition stabilized her condition changed -- rather than remaining in a bad state, Zahra has entered a new, stable, state. So, Frankfurt could conclude, Zahra has not been harmed by the change. Even if this move works in Zahra’s case, however, Frankfurt must say more to show that one can not benefit from something that keeps a bad state from getting worse since one may be lucky one’s state does not deteriorate.

Alternately, noting that Zahra would have been much better off without the disease, Frankfurt could advance the following conception of harm: Someone is harmed if and only if they would otherwise be in a much better state. It might follow that people whose bad state persists are harmed if Frankfurt also holds that those whose bad states persist are worse off than they would otherwise have been.

Neither contention is plausible. But even adopting the proposed conception of harm, for a moment, it should be clear that one is not necessarily worse off than one would otherwise have been if one’s bad state persists. Zahra, for instance, would have been in a worse state if her condition had not stabilized. (This is true on any plausible way of thinking about her state; she would have both lost her ability to feed herself and continued to degenerate if her condition had not stabilized.) Even right before Zahra got her degenerative disease she may have been worse off. She might, for instance, have had a virulent form of cancer removed just as she developed Parkinson’s.

The second problem for Frankfurt’s account is that people do not always need those things that allow them to avoid harm. Some harm is insignificant and people do not need to avoid insignificant harm. Zahra does not need to wear protective clothing even if this is the only way to keep her from getting minor bruises, for instance. Even if it is a law of nature that she will get minor bruises if she does not wear protective clothing, she does not need to wear such clothing.

Frankfurt might object that minor bruises are not harms because they are not severe enough to constitute harms. Alternately, he could say that Zahra does not need to wear protective clothing to avoid minor bruises because the clothing would be more harmful than the bruises.
Neither of these objections goes through. First, it is more plausible that the bruises are minor harms than that they are not harms at all. Second, it is hard to see how protective clothing is harmful. Perhaps Zahra would face harmful social stigma if she wears a full body covering. But, this is not clear. Zahra might live in Iran where she could fit in quite well with a padded burka that would protect her from minor bruises.

There are, of course, other ways Frankfurt might go here, but there are also other reasons to worry about Frankfurt’s account. Even undergoing significant harm can be beneficial. Enduring harm may be the only way to secure a greater benefit. Someone with cancer that can be successfully treated with chemotherapy may be harmed by the therapy but still needs it. Chemotherapy can cause kidney malfunction, blood clots, infections, and other serious problems (National Cancer Institute, 2007). When patients recover they may even be sicker than they were when their cancer was first discovered (though they may be better off than they would otherwise be). Even when the harms from chemotherapy are less severe than the harm the cancer would otherwise cause, they are still harms.

Frankfurt might argue that this is not a good case because chemotherapy does not harm but, rather, helps cancer patients. Without the chemotherapy patients often end up worse off than they would have been without the therapy. At least this seems right if one can only be harmed by something if one is made worse off than one would otherwise have been (Kagan, 1998).

While we did not challenge this conception of harm above, it is implausible. Suppose that Damien approaches Zahra as she is walking home. Damien attacks Zahra twisting her ankle in the process. Unbeknownst to Damien, a street on Zahra’s way home was covered with ice. If Damien had not attacked Zahra she would have slipped on the ice, hit her head, and become a paraplegic. But, because she was attacked Zahra called a cab instead of walking the rest of the way home. Damien has still harmed Zahra. People can be harmed even if they are not made worse off than they would otherwise have been.

Even if Frankfurt resists this conclusion or maintains a global theory of harm, there are problems with his account of needs. Frankfurt’s account unintuitively suggests that some people need things they do not (e.g. padded burkas). Nor can it account for the fact that some people need the things they do (e.g. stabilizing medication, if not chemotherapy).

Finally, consider a general worry about harm accounts. Some of the things people need they need to flourish, not merely to avoid harm. Intuitively, most children need education even if they will not be harmed by failing to get an education.¹ A good account of needs must be able to explain why this is so.

¹One might argue for this on the account of harm suggested above as follows: Some children’s lives might not be bad and even improve slightly over time if they do not receive an education, but even these children may need an education.
Social Role Accounts

The minimally good life account of need is not the only account that can explain why people need things that they do not need to avoid harm, however. David Braybrooke’s social role account provides one of the best alternative accounts of what people need to live a flourishing life. On his account, policy makers can determine what people need via a broadly consultative process. First, they must specify a list of necessary goods that enable people to fulfill four social roles – parent, worker, citizen, and housekeeper (Braybrooke, 1987). Then, policy makers must specify minimum standards of provisions for these things. The standards should be sufficient for each member of the population to carry out each social role. For, even those who choose not to occupy Braybrooke’s social roles will need many of the things that those who occupy the roles need (Braybrooke, 1987).

Unfortunately, some people do not need the things that would let them occupy Braybrooke’s social roles and others need things that they do not need to occupy these roles (especially if they hope to occupy other roles). A monk may not need to have children or be a worker but may need religious freedom. Braybrooke might not think that the monk needs religious freedom. But, even if he accepts this example, he might suggest that this freedom is just a part of the freedom of conscience necessary for the social roles at issue in his account. Braybrooke might maintain that people need freedom of conscience to be good citizens, for instance. On the other hand, Braybrooke might argue that the monk only needs the opportunity to have jobs and children. He might insist that people only need to have the opportunity to fulfill his social roles.

Perhaps some freedom of religion is necessary for the kind of freedom of conscience people need to be good citizens, workers, parents, and housekeepers. But, it is not clear that people need the kind of religious freedom the monk needs to fulfill these roles. Furthermore, Braybrooke cannot claim that people just need to have the opportunity to fulfill his social roles. It would be strange to say the monk needs the opportunity to have children. At least the monk does not need the opportunity to have children if he stays a monk and never wants the opportunity.

Braybrooke would probably respond to this last worry by saying that he is only concerned to give an account of what people typically need in a way that could be presented to the public. Most people need what will allow them to have children and work, though some people do fine without these things. Few people need the kind of religious freedom monks need. This, however, is just to say that Braybrooke is engaged in a different project than the one in which this paper is engaged, his account is not useful for our purposes. For, Braybrooke’s account does not fulfill the desideratum with which we started; it does not capture all the differences in individual needs because, intuitively, some needs

---

1Braybrooke says that not everyone will need to play every social role on his list and emphasizes that he is not concerned about idiosyncratic, or episodic needs. Everyone has a non-episodic need for food even if they are currently well fed. One has an episodic need for food when one is starving.
are not even rebuttably universal (i.e. few would think everyone needs the kind of religious freedom the monk needs). Braybrooke’s account, like harm accounts of needs, also suggests that some people need things that they do not (some do not need work or children but these are rebuttably universal needs). Although, like its competitors, the minimally good life account requires further cashing out, it has some advantages over the alternatives. We can, thus, conclude by turning to the connection between this account and human rights.

**Conclusion**

More work is necessary to cash out the accounts of autonomy and need suggested above. But, if the arguments in this paper go through, the accounts of need and autonomy we have sketched provide keys to avoiding the dilemma facing traditional theories of human rights. Recall that we suggested that many traditional theories either fail to account for the close connection between rights and autonomy or fail to account for the rights of all to meet their needs. That is, this paper has given us some reason to accept:

1. Human rights protect each individual’s ability to live a minimally good life.

For, we have argued for the following conclusions:

2. People need all and only what will enable them to live minimally good lives.
3. Autonomy is often necessary for and is partly constitutive of such a life.

So, if (1) is correct, human rights will protect each individual’s ability to secure autonomy and meet their needs. There is reason to take seriously a theory of human rights on which human rights, whatever else they do for people, protect peoples’ ability to live minimally good lives.\(^1\) Traditional theories must embrace something like (1)-(3) if they are to remain contenders or a new account of human rights might be necessary.\(^2\)

---

\(^1\)To account for the rights of those who cannot live minimally good lives one might specify that people should be able to come as close as possible to doing so.

\(^2\)The alternative accounts may have other advantages over the minimally good life account or there may be another option that merits consideration. Some, for instance, worry about autonomy and interest based accounts because they believe that these accounts are too individualistic. Rather, some suggest, we should accept a community based account of rights. I do not think that the minimally good life account need be viewed as a competitor to community based theories. Rather, I believe that they may each be useful lenses which emphasize different aspects of human rights. On this see (Wong, 2004).
Citations

I will start with three preliminary points. First, it should be noted that this contrast between Plato and Moore will rest on a certain reading of the two philosophers. Especially for Plato, as any other reading, that will be controversial. However, due to space limitations, I won't defend in detail this reading here. Rather, I will provide some indicative textual evidence supporting that certain reading and argue that, given this reading goes through, Plato and Moore based on analogous epistemological considerations rejected metaethical naturalistic reductionism and opted for metaethical nonnaturalism.

Second, there is a prima facie difficulty that undermines the very idea of building bridges between classical and contemporary philosophy. This prima facie difficulty concerns the fact that Plato and Moore were operating in disparate philosophical conceptual frameworks. Roughly, they were using different philosophical terminology. While Moore, more or less, is using the technical philosophical terminology of contemporary philosophy (‘property’, ‘naturalism’ etc.), Plato still struggles to lay down the foundations of western philosophy and has no such luxury at his disposal. He must invent technical philosophical terminology (something that perhaps was more effectively performed by Aristotle than Plato). This might give the impression that there is an unbridgeable asymmetry in terms of conceptual framework between classical and contemporary philosophy that inhibits any comparative approach between the two.

However, I think this difficulty is only superficial and can be set aside. Although Plato did not have (in equivalent Greek) such technical notions as ‘naturalism’, ‘nonnaturalism’ etc., he was still using other synonymous or analogous terms (in equivalent Greek) in order to address the very same philosophical problems that still haunt us. These other synonymous or analogous terms make his discussion translatable to the contemporary one (e.g. αἰθητή-perceptible and διάνοια-intellect, νοητή-intelligible etc.). Besides, Plato himself in the Sophistes (246a-c) makes clear that he is well aware of the great cleavage in metaphysics between naturalists and nonnaturalists. He calls the dispute between naturalists and nonnaturalists a gigantic battle (γιγαντομαχία). More generally, classical thought, though had no naturalism,
nonnaturalism etc. technical notions per se clearly had an understanding of what these notions involve, as the classical naturalist movement and the turn to naturalistic explanations insinuates¹.

Third, roughly, for the purposes of this paper, we can assume that Plato's notion of ‘form’ (ἐιδός or ἴδεα) is equivalent to the contemporary ‘property’². Also, we can stipulate that a natural property is a perceptible or empirical property. That is, a property with causal powers³. Accordingly, a nonnatural property is an imperceptible or non-empirical property. That is, a property that has no causal powers.

Now we can set out to explicate how the parallel between Plato and Moore is supposed to run. In the early ‘definitional’ dialogues, Plato’s Socratic character is searching for definitions of moral concepts⁴. Definitions that could provide us with knowledge of the moral principles that ‘ought to guide our lives’ (Rep. I 352d). He typically asks ‘What is Fness?’ where Fness stands for a moral concept. In *Euthyphro* asks for piety, in *Charmides* for temperance, in *Laches* for courage, in *Republic I* for justice, in *Hippias Major* for beauty, and in *Lysis* for friendship.

In the *Euthyphro*, the Socratic character seems to make explicit his theoretical assumptions about a correct quest for definitional knowledge of moral concepts (or virtues) (T.Irwin 2000: 147). He makes four distinct claims about ‘forms’ (or properties, in contemporary philosophical jargon) and what conditions an adequate answer to questions of the form ‘What is Fness?’ should fulfil:

1. There must be just one form of piety for all pious things (actions, people and so on) (5d1-5).
2. It must be usable as a paradigm that we can consult to say whether something is pious or not (6e3-6).
3. It must be describable without explicit reference to properties – just and unjust, fine and shameful, good and bad – that introduce disputes (7c10-d7).

¹See T.Irwin (1989), Chs.3-4 and (2005:51-8).
²Of course, unfortunately, these assumptions cannot be argued here. Also, the notion of 'form' is potentially misleading because Plato is famous for his theory of forms of the middle and late dialogues where forms are nonnatural properties. However, we shouldn't be misled to assume that when in the early dialogues Plato talks about forms (eide or ideai) refers to such nonnatural properties. In the early dialogues, as I shall argue, he seems to refer to natural properties. My suggestion is to understand the notion of form as equivalent to property and infer from the context whether Plato refers to natural or nonnatural properties.
³For such accounts of property see S.Shoemaker (2002), and D.Lewis (2002). For the natural-nonnatural distinction and the various ways that is being glossed in literature see M.Ridge (2003:2).
⁴These early ‘definitional’ dialogues are *Charmides*, *Euthyphro*, *Hippias Major*, *Laches*, *Lysis* and *Republic I* that is taken to be an independent dialogue that later on Plato used as the first book for the middle period Republic. For useful discussion on chronology see R.M.Dancy (2006: 5-11), C.C.W.Taylor (2000: 36-44). Note that the genuine character of *Hippias Major* is a matter of controversy.
4. The single ‘form’ (eidos) of piety is that ‘by which all pious things are pious’ (6d9).

The first and fourth claims are ontological while the second and fourth are epistemological. From claims one and four, Plato makes ‘the ontological assertion’ that there must be moral properties that our moral predicates purport to refer to and that these properties must be able to explain why all F things have the property of Fness. From claims three and two, Plato makes some epistemological assertions. From claim three, he asserts that these moral properties must be analysable, in pain of lack of informativeness and semantic confusion, non-circularly in non-disputed, non-moral terms.

From claim two, asserts that the suggested analysis must be usable as a paradigm that could resolve conflicts concerning whether something has or has not the property of Fness. This claim has important epistemological implications. It implies that the suggested analysis must be synonymous to Fness in order to be usable as a paradigm and resolve disagreement on whether something has the Fness property in question or not. This equally implies that the suggested analysis, in order to be synonymous and usable as a paradigm, must be reductive. Otherwise, the suggested analysis won't be usable as a paradigm. If we are going to have knowledge of the concept in question, we must have a reductive definition that is immune to counterexamples. Hence, the commitment to the (in)famous ‘priority of definition for knowledge acquisition’: knowledge of the meaning of a concept requires reductive definition of that concept.

Moreover, the epistemological claims two and three seem to assume implicitly that these moral properties are knowable because they describe what kind of semantic analysis of moral predicates could yield knowledge of these properties. By describing a correct method of semantic analysis, Plato seems to assume that in principle these properties are knowable. In addition, semantic analysis implies that these definitions should be naturalistic. As claim three suggests, moral concepts should be analysed non-circularly in other non-disputed, non-moral terms. Such non-disputed, non-moral terms it appears that should be natural properties. Properties that are relatively uncontroversial because they seem to impinge on our experience of the natural world. Therefore, Plato not only makes ‘the ontological assertion’ that there must be moral properties that our predicates purport to refer to and explain why all F things have the property of Fness (by claims one and four), but also implicitly makes ‘the epistemological assertion’ that these properties should be naturalistically knowable (by claims two and three). We should keep these two claims in our mind because the development of Plato’s thought will, crucially, refer to these claims.

Nevertheless, unfortunately, Plato’s quest for reductive naturalistic definitions of moral concepts in the early definitional dialogues always ends in

1For the commitment to the priority of definition for knowledge acquisition the Socratic character in the early dialogue has been charged for committing ‘the Socratic fallacy’. For discussion see Dancy (2006:35-47).
aporia, a dialectical impasse. Again and again, the Socratic character with the destructive elenchus cross-examines and refutes the suggested definitions of his interlocutors by providing counterexamples. Take for example Rep I and the attempt for semantic analysis of justice. Cephalus says that ‘justice is telling the truth and giving back what isn’t yours’ (331b). The Socratic character responds with a counterexample that refutes this definition. If you borrow a weapon, Socrates claims, from someone and then he loses his senses it seems wrong to give it back; nor is it right to tell the strict truth to a madman (331 c). Glaucon consents and then Polemarchus intervenes to give a more refined definition.

Polemarchus quotes a saying by the poet Simonides: ‘it is right to give every man his due’ and comments that this is an adequate definition. Under Socrates’ dialectical pressure Polemarchus restates his position as ‘justice is to help your friends and harm your enemies’ (334b) and adds that a man’s friends are those he thinks good, his enemies those he thinks bad(334c). Then Socrates, again, provides a counterexample. Sometimes, he says, we may by mistake hurt people that are good and our real friends but appear bad and help people that are bad and only appear to be our friends. Polemarchus finds this counterintuitive and accepts that there must be something mistaken with his suggested definition (334d). The elenctic argument continues with Polemarchus and then with Thrasymachus and finally at the end the Socratic character pronounces that “…we started off to inquire what justice is, but gave up before we had found the answer…so that I still know nothing after all our discussion’ (354b).

This plot recurs in all early definitional dialogues that constitute the great majority of all early dialogues. While these dialogues set out to find reductive, naturalistic analyses of moral predicates and have as their starting point a question of the form ‘What is Fness?’, they fail to reach a reductive analysis of moral concepts in natural, non-moral properties. In all these dialogues, the Socratic character provides counterexamples that refute his interlocutors’ suggested analyses and end in a dialectical impasse, reminding us the notorious Socratic disavowal of knowledge. As Socrates says at the end of Lysis(223b) ‘we have made ourselves ridiculous...we are friends of one another...but what a friend is, we have not yet succeeded in discovering’.

Plato’s Introduction of the Theory of Forms

We know from the transitional dialogue Meno that Plato was puzzled by the epistemological commitments of the early dialogues. With the Meno inquiry paradox and the invocation of the theory of recollection as a solution to the paradox, Plato rejected the commitment to 'the priority of definition for knowledge acquisition' and was thoroughly reworking his epistemology. However, the middle dialogues mark an even more radical transition in Plato’s thought. On the one hand, Plato introduces the notorious ‘theory of forms’ and expounds on its epistemological and metaphysical features. On the other hand,
Plato seems to neglect his search for definitions of moral concepts. This shift of focus from searching for definitional knowledge to the theory of forms and its ontological and epistemological aspects is not accidental. There are good reasons to think that these developments in Plato’s thought are interrelated.

I will focus on textual evidence from *Phaedo* and the *Republic*, though the pattern I will very briefly describe also appears to hold for the *Symposium* as well. In *Phaedo*, the Socratic character reaffirms ‘the ontological assertion’ but now also makes a step further (65d-e). He asks Simmias whether there is something just and something good and beautiful. When Simmias consents, Socrates asks him if he had ever seen such things. Simmias, of course, denies that he has ever seen them and Socrates goes on to claim that he hasn’t, because they can’t be grasped by perception but only by ‘intellect’ (διάνοια).

This is a very interesting innovation in Plato’s thought, because while he still endorses ‘the ontological assertion’, namely, that there are moral properties such as justice, goodness etc. that our corresponding concepts purport to refer to, denies that we can know them through perception, but instead claims that we can know them only through ‘intellect’. This creates a contrast with the early dialogues, where the Socratic character was trying to reductively analyse moral concepts to natural, perceptible properties e.g. whether giving back what isn’t yours is also just (Rep. I 331b) or whether courage is holding your line in battle and facing the enemy (La.190e). This seems to betray that Plato came to think that we can’t have epistemic access to moral properties through perception. If this is right, then Plato rejects the whole approach of naturalistic reductionism to moral concepts, because that rested on analysing moral concepts to natural, perceptible properties. But we shouldn’t rush to a hasty conclusion and wait to see how Plato’s arguments unfold.

Plato goes on to put forward ‘the argument from opposites’ for the theory of forms (74a-c). First of all, Plato, again, posits ‘the ontological assertion’. There must be a property of ‘equal itself’ or equality in virtue of which equal things are equal, says the Socratic character, and claims that this property must be different from particular things that have it (sticks, stones etc.) (‘non-identity claim’). Then he proceeds to explain why this property must be different from the particular things that have it. The reason has to do with perceptual knowledge again. He argues that though equal things sometimes appear unequal (think of the Muller-Lyer illusion) and vice versa, equality itself cannot appear as inequality and inequality as equality (‘self-predication claim’). What Plato means here, is that while the perceptible properties of being equal or unequal in respect with another thing are context-, perspective- and time-sensitive, the abstract, imperceptible property of equality itself cannot be unequal or inequality equal. For example, two sticks might be equal in length.

---

1 For discussion on the *Symposium* see Dancy (2006: 284-287).
2 Plato makes clear that the argument refers only to relational predicates that have opposites. Such predicates are moral e.g. good and bad, aesthetic e.g. beautiful and ugly and mathematical e.g. equal and unequal properties. The argument is not meant to apply to non-relational predicates like book, horse etc. However, Plato in *Republic* X (596a-597e) Plato uses the ‘one over many’ argument and generalizes for all predicates. See J. Anns (1981: 227- 240).
3 For discussion see N. White (2005: 280-294).
but not in width, they may seem equal now but not later, perhaps, because the one stick will be broken in two, they may seem equal in one’s perspective but not in another’s etc. Perceptible properties, hence, seem to suffer from, as is sometimes technically called, ‘compresence of opposites’ while the abstract properties, the universals themselves seem to be context-, perspective- and time- independent and, therefore, objective.

Plato, as the argument from opposites seems to indicate, thought that perceptual knowledge is unreliable. The perceptible property, action, state of affairs etc. which seems to be F in one context, perspective and time may be not F in another. Only the universal Fness itself seems to be exempt from ‘compresence of opposites’, though this is not grasped by perception but only by ‘intellect’. He indicates that if we want to keep the objectivity of moral discourse, namely, that we seem to refer to mind-independent properties, given that these properties seem to be imperceptible and, therefore, naturalistically unanalysable, we must postulate imperceptible, nonnatural ‘forms’. He proceeds to give an account of how these imperceptible forms explain why all Fs share the property of being F (96d-e).

He first gives an example of an unsatisfactory explanation. If we try to explain why X is larger than Y and say that this is so because is larger ‘by a head’ that would be a bad explanation. Plato claims that it would be a bad explanation because it suffers from 'compresence of opposites'. 'By a head' may make something larger or smaller respectively in different contexts and, hence, it cannot explain why large things have the property of largeness. It cannot explain why large things have the property of largeness because the same property in different contexts causes the instantiation of opposite states of affairs: being large and being small. Then the Socratic character proceeds to give what he calls a ‘safe’ explanation, an explanation that is ‘safe’ because avoids ‘compresence of opposites’ and gives an adequate, non-circular account(101a-e). He argues that all F things are F because of the form of Fness itself, not because of any perceptible properties that are susceptible to compresence of opposites. The form Fness is the cause that makes all F things F. As the Socratic character claims, beautiful things are beautiful because of the beautiful itself, not because of the perceptibles ‘having a bright colour or a shape or anything else of that kind’. The ‘argument from opposites’, the indictment of perceptual knowledge as unreliable and the resultant theory of forms help us see why definitions of moral concepts drop off the picture in the middle dialogues. If perceptible natural properties are susceptible to compresence of opposites, then we can’t find reductive naturalistic definitions of moral concepts. For an event, action, object etc. may be F in one context or time but not F in another and this

---

1 The term of art 'compresence of opposites' means that a perceptible property might be F (e.g. equal) in one context but not-F (e.g. unequal) in another.
2 Additionally, it is a bad explanation because is blatantly circular. It only restates that x is larger than y, but it doesn’t explain why is larger; what is the cause that makes it larger; though this is not Plato's own objection.
Plato, Moore and Ethical Nonnaturalism

destroys the project of finding natural, perceptible properties that could reductively analyse moral concepts. In fact, it has been argued that Plato seems to insinuate his abandonment of searching for reductive definitions when he says¹:

‘Now I do not yet understand…nor can I perceive those other ingenious causes. If anyone tells me that what makes a thing beautiful is its lovely colour, or its shape or anything else of the sort, I let all that go, for all those things confuse me, and I hold simply and plainly and perhaps foolishly to this, that nothing else makes it beautiful but the presence or communion (call it which you please) of absolute beauty, however it may have been gained;’(100c-d).

Although this passage is not crystal clear, it seems to me that in the light of the argument of the dialogue this is what Plato means. He says that analysing the concept of beautiful in terms of natural, perceptible properties like ‘its lovely colour’ confuses him, and prefers the ‘safe’ explanation of the nonnatural form. It confuses him because any perceptible, natural property would be beautiful or large etc. in one context, perspective or time but not in another and, hence, wouldn’t yield stable reductive definitions which he thinks are necessary for knowledge.

In the later dialogue Republic, Plato argues along the same lines but there things are perhaps more rectified. He again propounds the argument from opposites, criticizes the epistemological reliability of perception, postulates imperceptible forms and gives the ‘safe’ explanation why all F things share the property of being F (523a-524a). However, the Republic is clearer than Phaedo about his mistrust for definitions. When the Socratic character discusses the nature of goodness he rejects two definitions (505b-d). Firstly, he rejects the definition that virtue (or goodness) is knowledge because those who hold this view say that this knowledge is knowledge of the good and this is circular and uninformative. This criticism also appears in the transitional Euthydemus (292d-e) and is probably a criticism against the Socrates of the early dialogues, who held that virtue (or goodness) is knowledge of the good².

Secondly, as he also does in the transitional Gorgias (497a, 497d), he rejects the claim that goodness is pleasure –that is, the hedonistic account of Protagoras (355a-356a) - because there are some bad and some good pleasures. Since pleasure can be good in one context, circumstances etc. but not in another, it suffers from 'comprevence of opposites' and it cannot be goodness itself. Interestingly, when the Socratic character is being asked to give a definition of goodness he denies (506c). He adds: ‘…I am afraid it is beyond me, and if I try I shall only make a fool of myself and be laughed at’ (506d).

The Socratic character denies because thinks that if he gives any naturalistic definition, in virtue of the comprence of opposites, he won’t be able to avoid

counterexamples. Moreover, though he claims that goodness is a nonnatural property that we come to know not by perception but by some ‘intellectual seeing’ (484c-d, 500c), he does not give a reductive analysis of what this property is. Nevertheless, the Republic comes to reinforce the evidence of the Phaedo. Plato rejects the project of finding reductive naturalistic definitions of moral concepts because perception is susceptible to compresence of opposites and, thus, epistemologically unreliable. He postulates nonnatural imperceptible forms that can explain why all F things are F and defend the objectivity of moral discourse.

What does this mean in terms of the ontological and the epistemological assertions we met in Euthyphro? Since he believes that because of the unreliability of perception and the compresence of opposites, we can’t find reductive naturalistic definitions of moral concepts, he abandons ‘the epistemological assertion’ that moral properties are naturalistically knowable. He is still committed to ‘the ontological assertion’ and thinks that there must be moral properties that explain why all F concrete particulars share abstract universal Fness, but since we can’t have epistemic access to them through perception, these properties must be nonnatural and epistemic access to them should not be through perception but through some sort of ‘intellectual seeing’. Hence, Plato abandons ‘the epistemological assertion’ that moral properties are naturalistically knowable but he is still committed to ‘the ontological assertion’ that there must be such properties and concludes that they are unanalysable nonnatural properties.

Moore’s ‘Open Question Argument’ and Plato’s ‘Argument from Opposites’

Moore’s ‘open question argument’ in Principia Ethica is considered the starting point of contemporary metaethics. Moore (2000: 53 -72) with his ‘open question argument’ concluded that there are no natural moral properties. He claimed that it is always open for a competent user of language, without semantic confusion, to recognise that a state of affairs, event etc. has a natural property N e.g. pleasure or wealth, but still remain unconvinced whether the natural property N is synonymous with a moral property M e.g. goodness and, therefore, deny that the natural property in question is also moral.

Moore thought that this ‘open feel’ of natural properties showed that, if indeed there are moral properties, they cannot be natural. Rather, they must be nonnatural sui generis properties. If they were natural, moral concepts would be reductively analysable to natural properties and there would be no ‘open

---

Note that from Euthyphro (5d-6d) through Meno (70a-80d) to Phaedo (65d-e) and Republic (tacitly asserts it), Plato never questioned his ‘ontological assertion’, namely, that there must exist some sort of forms that explain why all F things share the property of Fness. N. White (2005: 302-303) rightly comments that this is ‘Plato’s fundamental insight’ and that ‘{t}he falsity of the insight would necessarily call his whole metaphysics into question’.

For discussion see M. Ridge (2003:3-5) and A. Miller (2003:12-25).
feel’ for competent users of moral language. Once the agent had discovered by a priori semantic analysis the natural property N that co-instantiates a moral property M, then she would reductively analyse the moral property and the question would appear closed. There would be no ‘open feel’. Moore concluded that all those philosophers that tried to define moral properties in terms of natural properties were committing ‘the naturalistic fallacy’.

The question now is how Plato and Moore run in parallel. Recall that Plato rejected ‘the epistemological assertion’ of the earlier dialogues that moral properties are naturally knowable because, according to the argument from opposites, natural, perceptible properties are susceptible to compresence of opposites. From there he carried on to defend his ‘ontological assertion’ and, like Moore, stated that in order to save the objectivity of moral discourse, we must posit nonnatural properties that are immune to compresence of opposites. It should be obvious by now that this is what Moore also did. Moore’s ‘open question argument’ relies on the same epistemological assertion namely that, if there are natural moral properties, then these must be naturalistically knowable by a priori semantic analysis. Semantic analysis implies that we can have epistemic access through perception to these properties and be able to reduce them in natural, non-moral features of the world. Once we do this, the ‘open feel’ should evaporate.

However, Moore complains, any suggested naturalistic analysis of moral concepts is doomed to meet this ‘open feel’. Every time that a possible analysis is suggested, the ‘open feel’ allows people to remain unconvinced whether a natural property is also moral and provide counterexamples in a Socratic style that refute the definition. Moore took the ‘open feel’ intuitions we have about naturalistic analyses of moral concepts to show that there are no natural moral properties.

Moore's ‘open feel’ complaint is analogous to Plato's complaint that natural, perceptible properties are always suffering from ‘compresence of opposites’. Pleasure might appear good from one perspective or context etc. but not from another and, hence, it will be questionable whether the natural property of pleasure is also synonymous with a moral property like goodness. This is why the Socratic character of the early dialogues takes advantage of this ‘open feel’ of context- or time- or perspective- sensitivity of suggested naturalistic definitions in order to provide counterexamples and refute them.

**Conclusion**

I have argued that both Plato and Moore rejected metaethical naturalistic reductionism because of analogous epistemological considerations. Plato in the middle dialogues repudiates ‘the epistemological assertion’ of the early dialogues, namely, that moral properties are naturally knowable because natural, perceptible properties suffer from ‘compresence of opposites’. Moore

---

1Actually ‘the naturalistic fallacy’ is misnamed because it is not a logical fallacy. For discussion see W.Frankena (1939: 464-477), M.Ridge (2003: 1-5).
also believed that the ‘open feel’ of moral concepts shows that they are not naturalistically analysable and, hence, they are naturalistically unknowable. He concluded that there no natural moral properties. Plato’s ‘compresence of opposites’ and Moore’s ‘open feel’ are analogous and stem from the common epistemological worry that we seem to have no perceptual access to moral properties in nature. If we have no perceptual access to moral properties, then we can’t analyse the corresponding moral concepts in naturalistic terms. However, both took as fundamental ‘the ontological assertion’ and moved on to metaethical nonnaturalism.

Bibliography


(1989). *Classical Thought*. OUP.


Translated by W.R.M.Lamb.


he author of one of the most complex and sound social theories of the medieval period is the Central Asian philosopher and naturalist Abu-Nasr Al Farabi, who lived in IX-X centuries. In his social and ethical treatises he considered the following topics connected with origins, essence and forms of state governance.

1) Origins of the structure and forms of social associations
2) The city–state, features and life of the municipal (state)
3) The state and the forms of its governance
4) The place and responsibilities of human beings toward society.
5) Happiness as a final goal of social association, ways and methods of its achievement.

At the centre of al-Farabi social doctrine was the problem of interaction between humans and society, their rights and responsibilities towards each other.

Al-Farabi wrote ‘one of the main tasks of civic science is a description of the moral and virtuous activity of people, which leads them to true happiness, and how a political or state association could help them to achieve it. The primary function of a political leader has to be the foundation and reservation of morality. The key paths for the attainment of true happiness are virtue, chastity, and generosity. Any activity that is opposite to them is evil and imperfect. Only the optimal social environment is able to develop positive human action’.1

Al-Farabi divided civic science (science about the art of governance) into two interrelated parts. The first one aims to explain and substantiate the essence of happiness as the main goal of state governance. He distinguished between true and fictitious happiness. In addition, this section of science considers the actions behaviour, morals and manners of cities and peoples and divides them into virtuous and non-virtuous. Another section emphasizes the social associations of people and ethical norms, which could be used as a

---

foundation for the association and behaviour of people. In this sense his ideas are in tune with Kant’s thoughts about ‘civil society as a house for the whole of humankind’, built on moral foundations and whose main purpose is the happiness of each person within a society.

Al-Farabi suggested his own, very original idea of the parentage of society. ‘Every human by his nature for the maintenance of his life and for the achievement of perfection needs many things, which he is not able to provide himself. In that case he needs some association of people, where everyone provides certain things, which are necessary not only for themselves, but also for others and which they are able to share. And because of that, within an association whom each helps the other, where everyone shares some part of his product, the human being is able to achieve perfection’.1

This position has a lot of similarities with the theories of Western philosophers of XVII-XVIII centuries, particularly with Rousseau and his doctrine about “natural agreement”, where the origins of the society are interpreted in the same way.

In the structure of social organization, al-Farabi considered great, medium-size and small society. Great society is the unity of all people, all nations and ethnical groups of the Earth; medium society – is a society of a particular ethnical group or nation, small society that of a city. The city is a first step toward perfect society, unity of cities constitutes the nation, and the association of nations constitutes the whole of humankind.

Al-Farabi drew a sharp distinction between the virtuous or ideal city and the ignorant one. The virtuous society (city) he defined as ‘the city, where associations of people have the aim of mutual help for the achievement of true happiness. The nation, which consists of many cities, where people help each other to achieve happiness, is a virtuous nation. Therefore, all of our Earth would be virtuous, if nations would help each other to achieve true happiness’.2

Al-Farabi compares the virtuous city-state with the human body, where all organs and systems are interconnected and help each other to keep life and make it full. The same situation should obtain in the city, where all members depend upon each other and implement different, but complementary functions.

The head of such ideal city is supposed to be virtuous and respected by every person. Its main aims are the creation of conditions for the achievement of true happiness for all members of society. In that city goodness and justice would predominate, and injustice and evil would be subdued.

Ignorant cities have no idea about true happiness and never try to achieve it. They care only about their physical health, wealth and enjoyment. The ignorant city comes in many varieties. One such city, where people care only about necessary material and natural needs, such as food, beverage, clothing, habitation and so on, and do not care about spiritual improvement. Another one is ‘a city of illusion’, where the main aim of people is the achievement of wealth and prosperity. The city of ‘meanness and unhappiness’ provides only

---

bodily enjoyment. ‘The ambitious city’, where citizens dream only about honour and greatness; ‘the power-lover’ city, where only encourages thought about the conquest of others. In the ‘voluptuous city’, people do not want recognition and power, and there is complete anarchy. In the ‘changeable city’ people try do good things, but they accept wrong ideas. ‘The stray city’, has inhabitants who make violent statements about perfection and happiness.

Al-Farabi thought, that in the virtuous city there exists a very complicated hierarchy of classes, because ‘everyone is doing his own businesses’. The basis for such a hierarchy is the talent of each particular human being, first of all with intellectual abilities, and also knowledge acquired in the process of living experience and scientific learning.

The inner task of the state, according to al-Farabi, is the implementation of necessary measures, which could provide true happiness for its inhabitants, establishment of justice, the enlightenment of people, moral and intellectual education, and cultivation of higher dispositions, all of contribute to the achievement of happiness.

Solving all the other problems economic, political, etc.-depends upon the spiritual health of a nation and the degree of its spiritual improvement. Al-Farabi thought that all religious confessions are equal and the extent of spiritual improvement of any nation depends upon its religious persuasions.

One of the most important topics, which had been considered by Al-Farabi is the personal characteristics of the head of the virtuous city. He thought that nobody is able to unite all the best characteristics of personality. In that case, he suggested not one, but two or even more equal people who would be able to combine different virtues and together rule the state. But if it is possible to find one person, who would unite at least the majority of the best qualities, better to give him the power.

In his social and political doctrines, al-Farabi writes about democracy. He thought, that among all of ignorant cities, the nearest to the virtuous city is the voluptuous city, because citizens of that city are absolutely equal. In the voluptuous city there are a lot of different inhabitants, some of them virtuous, some of them voluptuous. Probably, they would esteem a governor who is a virtuous respectable person. The voluptuous city allows virtues people to achieve leading positions and create an ideal city.

Al-Farabi considered freedom and equality as very important values of social life, but at the same time he thought that without strong state government, which is able to suppress anarchy and create conditions for the development of virtuous qualities it is impossible to achieve true happiness.

Al-Farabi emphasized three forms of governance, which can be defined as monarchy, aristocracy and feudal-democracy. Al- Farabi thought that all three forms are equal and all of them have the main goal of enlightenment, and the creation of conditions for happiness.

This idea of al-Farabi about the three forms of governance has many parallels with the theory of Montesquieu concerning forms of governance.

---

Although al-Farabi lived one thousand years before above mentioned Western Philosophers, one can enumerate quite a few parallels of this kind.

**Political ideas of Beruni, Ibn Sina and Navoi**

First-rate thinker and naturalist of the Medieval East is Abu Rihan al-Beruni. In his social and political doctrines he drew on a comparative analysis of the cultures of different nations, for instance Indian, Greek and Arabian. He thought that the best elements of all cultures included humanistic values. He also saw no principal differences between the philosophical foundations of Hinduism and Islam. All religions are equal and the Koran could exist harmoniously together with other Holy books. Humanistic values are universal. Thus the people, who are carriers of those values, also should be equal.

The social and political theories of Ibn Sina (Avicenna), who was widely recognized as the greatest Muslim philosopher, were formulated on the basis of his humanistic understanding of the place of the person in the world. He thought that feudal society is too conservative does not allow enough freedom for the person. A better would be one society where education and science would have a leading position.

Another requirement for a better society is a close connection and interaction between nations and states. Ibn Sina wrote, that ‘a human being cannot satisfy his personal requirements without communication with other people... Therefore, it is necessary to make agreement between people and communities’.

Some thoughts on the theory of civil society were expressed in social and political doctrine of the great Central Asian poet and humanist Alisher Navoi (XIV- XV centuries). The main social and ethical ideals were stated in his works of art and poems ‘Mollification of Righteous, ‘Beloved of Hearts’ and ‘The wall of Alexander’. Navoi wrote that the ideal of social life is harmony between different classes. This ideal could be reached by social reformation, which would be led by a just and wise governor. According to Navoi, justice is the main social virtue.

Social relations are based on the different, often conflicting interests and goals of many individuals, on difference of opinions and desires. In these there is a source for contradictions and conflicts. The main aim of the governor is hence to lead people to mutual consent and harmonious relationships. The mission of the governor is to keep the peace and defend the weak from their potential oppressors. The basis for ideal society should be the primacy of moral and ethical norms over political ones. Elite classes have to be positive

---


examples of justice and virtue. The style life of governors and they are daily activity have to be a model for public.

Very important, in that case, are personal qualities of the leader. He has to personify high moral qualities and practice justice and humanism. Some of the qualities, such as modesty, kindness and generosity have special importance for the governor. Who ever unite all these characteristics, is an ideal man. Justice as a social ideal has to be realized in three respects: moral, political and legislative.

Morally, justice is interpreted as charity and characterizes interpersonal relationships. As a principle of moral life, justice means the requirement to do for another the same things, that you want done for yourself. For governors and their administrations is their moral duty to care about the public and defend its interests, especially the poor and helpless. Authority and the power of the king have to be based on justice.

Politically, Navoi considered justice as a state virtue, providing the viability of the ideal society and serving as the moral basis for social and political life and governance. Justice is also a characteristic of social relations and the whole social order. It is a fundamental principle and criterion of government. A measure of justice is reward for virtue and punishment for vices. It has to be the main way of the governor’s policy. Humanism has to be the main philosophy of administration. In the poem ‘Mollification of the Righteous’ he wrote that the king is not the personification of oneness, all of the people are slaves of the one Allah. Some governors in their personal characteristics yield to nations, but their important position and wealth oblige them to do good and merciful things. It is necessary to suppress despotism of iniquitous governors, because it is a requirement of justice, which is not just an ethical, but social and political category, where all three components complement each other.

Legally, justice means order and law and it is a guarantee of the security and defence of every member of society. The governor has to keep vigilant watch on the observance of lawfulness. The governor has to be accountable to the law as well as to his nations. He can not decide anything without the advice of wise councillors. Otherwise another governor should be elected. These ideas were similar to ideas of constitutional monarchy, which were developed by European philosophers of XVII-XVIII centuries.

Navoi was always against the revolutionary transformation of society. On the contrary, he thought that social problems could be solved by self-improvement of a human being, his or her inner development. Society has to offer individuals a maximal spectrum of rights and freedoms and support the spiritual progress and inner development of a human being.

Political and Philosophical Ideas of Jadidism

Jadidism movement formulated modern socio-philosophical ideas in Turkistan. It was a movement of Enlightenment in Central Asia. The movement derived its name from its advocacy of the usul-i jadid, the new
method of teaching Arabic alphabet to children in the maktab, the ubiquitous elementary school of sedentary Muslim societies of the region. Implicit in the concept was a new way of looking at the world. The development of industry stimulated an inflow of high by qualified specialists to this region. The most important item in the lexicon of Central Asian Jadidism was *taraqqiy*, a term that covered the notion of progress, development, rise and growth. The Jadid’s assimilation of the idea of progress, the notion of history as open-ended change, altered the way in which they saw the world and their place within it. If progress were a fact of nature, then it was incumbent upon Muslims as a community to strive for it too. Progress and civilization were accessible to all; the only precondition was the cultivation of knowledge. Knowledge also explained for the Jadids the superiority of the “more advanced” societies of Russia (and Europe in general) over Muslim society. These societies were living examples of the links between knowledge, wealth, and military might that the Jadids constantly asserted. This in turn have influence on the education system in Turkestan and finally on social relationships.

Before, life in Turkestan was based on patriarchal relations. It had a positive side, because society in general developed without social conflicts and tried to maintain tradition in all spheres of social life. But social conservatism created huge problems as well. The public did not have the opportunity to express its opinion, to discuss social challenges. The secular educational sphere was not developed either. Religious schools (medrese) were concentrated on teaching the Koran and other holy books of Islam and never prepared specialists for industry, agriculture and so on. It is interesting, that tsarist Russia tried to perpetuate this order. Russian administrators in Turkestan wrote in their reports, that the youth of that region is very well behaved and is not as inclined to freethinking as the youth in Russia.

Under those conditions the ideology of Judidism arose. Jadids suggested the reform of the whole social sphere. A. Khalid wrote ‘like in Europe, the enlightenment movement in Turkestan was a movement against medieval feudal relationships, remnants and traditionalism. It was a struggle of progressive forces, which tried to take the country beyond feudal relations’.

According to Jadidish, social reforms have to be based on the traditional identity of the people and acknowledged by the Koran. They began with educational sphere. Many of them opened private schools, where they tried to combine religious with secular education. Much more attention was devoted to secular (hard and economic) sciences the in old medreses.

One of the most significant representatives of the Jadidism movement was *Makhmudjon Bekhbudi* (1875 - 1919). In his activity he popularized secular science and culture. In 1913 he started to publish the ‘Samarqand’ newspaper and also ‘Oy’na’ (Monthly) magazine, which existed up to 1915. Bekhbudi was the author of several articles about life in Central Asian nations and also he wrote ‘Alifbe’ (Alphabet), ‘Geography’ and many other textbooks for Jadidist schools. In his political doctrine he suggested the establishment religious

---

autonomy. His project reflected the aspiration of trade classes in Turkestan to develop a national culture and economy and to have local representatives in the governmental administration. Ideas of democratic reforms of a society were also supported by Judidist organizations, such as Young Khiva and Young Bukhara. These organizations had a liberal wing, members of which tried to reform social relations on the basis of democracy, and radical wing, representatives of which took part in the revolution of 1920.

Sattarkhan, Anbar Atin, Zavki, Mirmykhsim, Avlony, Nozir Turakulov and others proposed the most interesting ideas, which were propagated in religious form. One of their most interesting statements was that Islam does not contradict the idea of the progress. Islamic priests and men of religion played one of the most important roles in the Jadidist movement. One of them, Fitrat, talked about religious tolerance. 'There is no doubt, that people worship different religious, live in different countries and belong to different ethnical groups, but all of them are human beings. In other words, they are brothers and sisters. Therefore all of them have to respect the dignity of each other’ and have to establish an era of Universal Brotherhood. Thus there would be a new epoch of happiness, peace and universal prosperity1'. Sattarkhan Abdugafurov emphasized that the Koran, which was translated into English, French and Russian still is a Holy Book. That is why it is very important to translate this book into modern Uzbek. Furkat was convinced that society could be improved only when science and education would be developed, and the government of the country would be led by highly educated, wise ruler. He wrote that the 'learning of different sciences is necessary for a human being’s benefit and for the benefit of Muslims'2.

Jadidy substantiated necessity of education for women. In 1906 in Urgench one of them – Khusan Dushaev open the first school for girls. In other cities of Turkistan girl-schools were also opened, where mostly secular sciences were taught. Such outstanding jaddid scholars as Avlony, Rasuly, Fitrit, Bekhbuyd and others worked as schoolteachers. They wrote and published the textbooks for their schools and promoted equality between men and women.

The main idea that united the most outstanding thinkers in Turkistan was oriented towards spiritual freedom and human dignity. Rather then oppose one class or group against another; they believed that all human beings are equal irrespective of their religion, social origins and so on.

Ahmad Danish, who lived in Bukhara, saw that one way to achieve reform, was to make secular education, especially scientific knowledge more widely available to masses. S. Abdugafarov agreed with Danish and added that human being as the most evolved creatures are endowed with reason and so naturally seek self-understanding. In this light, it is important for human to find out where they come from, who they are, and where they are going, as well as the reasons for our happiness or misery. The key to answering those pivotal

questions is knowledge. Knowledge is also important to help distinguish between truth and falsehood. However, knowledge is only an instrument in the surviving of humanity. It should not regard or amend in itself. A very important role for knowledge lies in its promotion of secular thought which Danish, Adugaffarov and others believed should act in conjunction with religious reforms to being about the desired social changes.

Jadids tried to an independent democratic state in Turkistan. For example, Bekhbudi founded the People Muslim Union in 1905. The objectives were religion autonomy and self-administration of Muslims in Turkistan. In 1914, he came up with his own proposal for administrative reform. He advocated the creation of an administration that would include all ethnic groups in Turkistan. The main office would be in Tashkent, with departments in other regional centres. Each department would have four administrators, all of whom would require to have at least completed their secondary education, and preferably their high education. The administration would encourage secular and religion officials to work together, in conjunction with the mosques, secular schools, religious medreses, and other social sectors. In the meantime, this sector would be reformed in tune with Bekhbydi’s vision. Accordance with Behbudi’s ideas about the democratization of the political administration of the various social sectors had followers such as Munnavar Kari and others. In 1917, they founded an organization that they called ‘Shari’a Islamiya’. The main idea was the creation of a special from of administration in Tashkent and in spirit of February Revolution in Russia.

Bekhbudi and Kari both represented the political centre and the reforms minded current within Jadidism. The radical current was represented by Mukimi and Zavki. They suggested revolutions act on social and economic relationships within society have to be destroyed in order to achieve a democratic egalitarianism. In addition to the Russian Tsarists influence, local feudalism was also blamed for the corruption, iniquity, political opposition, and generally backward of Turkistan’s society. Initially, they had high hopes of converting the educated to their cause. When the Jadid turned materialistic, they dismissed the whole lot of them as bourgeois reactionaries who must therefore be eliminated. The goal was still an independent democratic Turkistan, but to achieve by means of revolutionary action.

Concluding Remarks

There is no doubt that may of the foregoing ideas were also expressed in the work of thinkers in countries outside of Central Asia. This is not only tone in areas of the Islamic world, such as Turkey or Egypt, but in other civilizations well, for example in Central European and even Latin America. The clamour for reform was universal at the beginning of the twentieth century. We often find reference in all those writings to the value of the person, equality before the law, the importance of education and the pursuit of knowledge, and the formative role played by religion and spiritual life. However, what is of special
instinct is for us to note the degree to which such modern thinking could be found among Central Asian leading intellectuals, its relevance for the quest to read an accommodation between Islam and modernity, its resonance with the contemporary concern with civil society (our main focus here), and its possible contribution towards more harmonious relationship between East and West at a time when many are unfortunately using the language of clash of civilization.

**Bibliography**

Selected works of Central Asian and Middle East thinkers (1967), Moscow: Progress.
The Abolition of War: A Philosophical Imperative

Joachim L. Oberst

In this paper I shall develop a response to the universal acceptance of war as a means of peace.¹ At the end of World War II, shocked at its own devastation, humanity made preparations for the containment of war. The fascist perpetrators of war, Germany and Japan, abolished their military traditions, and the international community placed itself under the rule of international law within the body of the United Nations. Despite these concessions, the end of World War II had already triggered the Cold War, most notably with the dropping of the Uranium bomb over Hiroshima on August 6, 1945, and the Plutonium bomb over Nagasaki three days later. At the dawn of the Atomic Age the pre-emptive strike by the US in Japan aimed at the Soviet Union carried the message of deterrence as the only way of providing peace. The Soviet Union understood the message and became nuclear.

Forty-five years later with the end of the Cold War, the world missed the opportunity to establish a peace no longer built on hostile armies. With the imbalance of power following the dissolution of the Soviet Union, preemption officially became the logical extension of deterrence particularly for the only Superpower left able to employ the measure successfully. This was met with the ironic, since counterproductive, consequence of an even heavier reliance on deterrence by the potential victims of preemption. The nuclear weapon is the obvious, since most apt response to the seemingly unchallengeable threat of preemption. It is counterproductive as a response in two unintended antagonistic ways. On the one hand, preemption aims to undo the threat of deterrence with an incapacitating first strike. The capacity of sudden and instant disarmament and defeat of any opponent intends to persuade potential enemies never to challenge the absolute Superpower with military means. However, total devastation by the nuclear weapon bears the promise of a credible deterrence, which is why such persuasion will remain unsuccessful. It will drive vulnerable nations into atomic statehood. On the other hand, it is precisely this promise of credibility intrinsic to the nuclear weapon that will tempt the Superpower into a preemptive strike. This development has thrown

¹Indicative of the universal acceptance of this doctrine is the following statement by George W. Bush in an address to the United Nations General Assembly on September 21, 2004: ‘For the sake of peace, there must be serious consequences, [i.e. war].’ http://www.whitehouse.gov/news/releases/2004/09/20040921-3.html
the world into the vicious circle of an uncontrollable arms race, fueled by the nuclear weapon used both for the deadly first strike policy of preemption and as a defense against it. This vicious circle is the historical reality of today.

Intrinsic to this development is the problem of perception. The mere perception of a threat justifies a first strike as anticipatory self-defense within the framework of the just war theory. In the strict and best sense perception bears as psychological judgment its own justification. In the worst, political sense as deliberate fabrication it is a pretext for aggression. In either case it blurs the line between fact and fantasy. To escape the traps of the vicious circle, not just a political, but a psychological analysis is needed. Here lies the need for philosophy. Philosophy can analyze politics in terms of psychology. In what follows I shall consult Immanuel Kant and Friedrich Nietzsche for an anamnesis of the human condition. This will be followed by a reflection on Martin Heidegger and Søren Kierkegaard for possible remedies. With the assistance of all four I shall demonstrate that the only promising response to the human predicament of war is the uncompromising and unconditional implementation of the philosophical imperative of non-violence. To do so I will first consider Carl von Clausewitz’s philosophical account of war laid out in his book On War from 1832.¹

Clausewitz’s Conception of War

The posthumous publication of Clausewitz’s Opus Magnum On War by his wife a year after his death in 1831 is the unfinished product of some twelve years of relentless work. The book is composed of three parts divided into eight books. The last two books are unfinished sketches.

Clausewitz compares war with a wrestling match. The fight between two wrestlers illustrates the nature of war. War is fought with a force that is both responsive and aggressive, thus never self-contained. Its nature is to be unrestricted. War aims at self-supercession toward its ultimate goal: victory of one side and defeat of the other. Without this final result war is not over, even if it is suspended with a truce. The essence of war consists in its teleological goal of a final result for which the utmost force must be used whenever necessary and wherever possible. The reciprocal nature of the extreme triggers the spiral of violence. There is no escape out of this vicious circle. ‘War’, says Clausewitz, ‘is an act of violence pushed to its utmost extreme’ (103/14/85). The extreme is determined by the means of force-maximization: intelligence, strategy, technology and the morale an ideology can foster. These are the roots of war’s self-perpetuation. No final result is ever absolute. Defeat itself can become the casus belli for a revenge war. It lies in the nature of war not to

leave any room for ‘real’ or ‘perpetual peace’. Reliance on war foregoes the prospect of peace. We shall see that it is here where the student of Kant lacks Kant’s prophetic vision and parts company with him.

Entrenched in the time of European warfare Clausewitz is a pragmatic realist. He cannot foresee a warless future. War is as much a part of human history as anything else that constitutes the human reality. If war is an inextricable part of human life, politics cannot exclude war from its domain. War is part of politics. Clausewitz conceives of a highly expedient relationship between politics and war. He defends his position against any categorical exclusion of war from political affairs. Clausewitz is very adamant about this. War is never the end of politics, but itself a political act and tool designed to achieve political goals. Indeed, he writes, ‘war is a mere continuation of policy by other means’ (119/22/99). The assertion appears to entail its own refutation. Clausewitz acknowledges the categorical shift that comes with the introduction of war into the political sphere. The ‘other means’ that constitutes the act of war stands in contrast to non-violent political commitments in domestic and international affairs. Whether we qualify this paradigmatic shift as an end of politics or its extension seems to be a semantic exercise. Even Clausewitz can see that with war something different has entered the political stage. However, Clausewitz integrates the novelty of war into the political landscape to provide it with a distinct place. War is a unique ‘language’ that politicians are compelled to speak. The fact that this language is universally understood demonstrates that although it has ‘a grammar of its own, its logic is not peculiar to itself’ (402/217/731). Everyone shares it and thinks along its lines even if the grammar, its peculiar mode of conduct, appears difficult to adopt. Clausewitz contends that this difficulty disappears even for ‘the novice to the battlefield’ within half an hour of combat experience (159-160/46-47/132f).

Hence, Clausewitz concludes, the exclusive nature of war has its normalcy in human affairs and offers itself as an instrument to achieve political goals.

The National Security Strategy Document (NSS) published by the White House in 2002 is a clear application of the Clausewitz doctrine. If war is essential to the human condition one has to make the best of it. ‘Different circumstances’, the NSS states with the words of George Bush (West Point, New York, June 1, 2002), ‘require different methods, but not different moralities’ (3). Ironically, the different methods are always the same, an improved readiness for war, which culminates in preemption, because the ‘circumstances’ are ‘different’ today: deterrence as a means to secure peace has failed. This failure is intrinsic to deterrence. In today’s technological arms race, deterrence has become outdated as a sufficient defense mechanism. The ‘moralities’ have not changed. This is a time of war, more so than ever. The ‘war on terror’ is a global and indefinite war. Its omnipresence has suspensions, but no end is in sight until so-called ultimate defeat and victory is achieved. Such prospect is grim on this totalizing scale. War has become

1 http://www.whitehouse.gov/nsc/nssall.html
perpetual. It is no longer designed to end war, but to make peace obsolete.\(^1\) Clausewitz himself did not believe in a theory of deterrence. Armed forces aim at subjugation to assert supremacy. The peace that follows is an enforced peace. Such a peace can only be temporary or rest on the annihilation of the opponent. This realization that preemption is a logical extension of deterrence is Clausewitz’s contribution to our time.

In summary we can conclude: The nationalistic drive in human history towards the Superpower Status has confirmed Clausewitz’s philosophical analysis of the nature of war.

(1) It is intrinsic to nations to be in preparation for war, thereby seeking to outdo each other’s war efforts. The theory of deterrence has made manifest that the arms race both sustains and threatens the equilibrium of an armed peace to the point of producing ever more powerful weapon-systems. The sky (space) literally is the limit of war. There is no limit to belligerent solutions to conflicts. Warmongering with or without actual warfare spins the cycle of violence into self-destructive proportions yet largely unknown to humanity.

(2) War can never guarantee long-term solutions to any conflict. Suspension of war in ceasefire only postpones it. Even defeat and victory are only intermediate goals, never absolute. They set the stage for new wars. War conditions the human reality.

(3) Clausewitz’s concept of war stands in agreement with Hobbes’s notion of a bellicose mankind (\textit{homo homini lupus}) in need of being tamed into voluntary submission to the political authority of legal, punitive violence threatened by the sovereign ruler in cases of disobedience or civic violations. The civil state of nations is at best one of truce. For Clausewitz it is cosmetic rhetoric to reject his dictum that ‘war is the continuation of policy by other means.’ To call such ‘continuation’ the ‘end’, ‘replacement’ or ‘failure’ of politics, when it is \textit{de facto} a political tool of choice by politicians to achieve ego-national goals is only of semantic-technical, not philosophic-political significance.

\section*{Kantian Limitations on the Absoluteness of War}

Clausewitz may have been a faithful student of Kant’s philosophy, but he did not share his foresight. Kant undermines the moral-philosophical validity of the Clausewitzian war theory. In his piece \textit{To Perpetual Peace} from 1795 Kant outlines in detail pragmatic steps towards what Nietzsche calls ‘real’ peace, and what Kant, aware of its reception as utopian, provocatively calls ‘perpetual’ peace.\(^2\) This designation is not without intended irony, as is clear from the opening lines. Politicians, Kant says, ‘can never get enough of war.’


Philosophers are for this reason both compelled and reduced to ‘dream[ing] that sweet dream’ of ‘perpetual peace’ indefinitely. However, this utopia is never out of reach as long as the dream is dreamt in all its moral and practical consequences.

Nations at war must never settle for less than peace. In order to be able to emerge out of their war-faring habits, they must first make sure never to shut the door to the possibility of peace by the ways they conduct war. A fundamental trust in each other’s humanity must be preserved even during actual warfare. This condition puts serious restrictions on the conduct of war. Kant’s formulation of such restrictions reveals how far humanity has regressed since his time. For the sake of brevity let us simply state that they prohibit all breaches of promises, the mistreatment of prisoners, the use of torture, the targeting of civilians, assassinations and spying. Such violations undermine all prospects of peace. Continued reliance on ‘such acts of war’, as Kant defines them, will in the end plunge humanity into the ultimate ‘war of extermination’. Although Kant’s time was not within the reach of its own annihilation, Kant foresaw in the persistent reliance on war the end of human history as a real possibility.

With the end of WWII humanity has reached the fatal crossroad. Humans have since been standing under the dictum of a pressing historic alternative. Either humanity permits perpetual peace to open the door to a flourishing future, or perpetual peace will, as Kant famously put it, ‘occur only in the vast graveyard of humanity.’ Kant foreshadows the famous words of Martin Luther King. King repeats these words as a warning in his final speeches. ‘Nonviolent coexistence or violent coannihilation’ is the choice humans still have, he says almost optimistically. However, he admonishes, the choice is no longer just one ‘between violence and nonviolence’. It’s ‘nonviolence or nonexistence’.

Kant advocates King’s position some 170 years earlier. This explains his categorical juxtaposition of his concept of peace to truce. Kant rejects truce as a form of war. A truce must never take the place of peace, since truce is based on continued hostilities, even if actual warfare is temporarily suspended, whereas a peace treaty abolishes all causes for war. The abolition of all causae belli, of all reservations and ill will between warring nations, will make manifest that perpetuity is intrinsic to peace. This recognition reveals the pleonasm in the title Kant has chosen for his philosophical sketch. Were this recognition common sense one would not have ‘to modify [peace] by “perpetual”’, since such modification obviously ‘smacks of pleonasm’. Its necessity is a symptom of our time, which not only lacks the true concept of peace, but also its fundamental experience. The consequences are as far-reaching as they are obvious. Peace in its truest sense entails the abolition of standing armies. For in their design and purpose armies constitute a threat

---

externally to other nations, and internally to the human being who as a professional soldier is reduced to being a non-autonomous tool in the hands of the nation’s rulers. Such reification is inhumane. It violates the fundamental ‘rights of humanity’ to self-determination by turning the soldier into an amoral and immoral killing machine thereby depriving her and him of their humanity, as something dispensable to be wasted at will.

It is clear that such reasoning rejects deterrence as a viable means of peace-making. Deterrence throws humanity back into the state of nature as hostilities are being bred in legal loopholes. The lack of legal authorities to which all nations submit in order to transcend from the natural state of war to a civil state of peace turns the might of universal rights into a might of ego-national rights. Raw might, however, stands under the command of being dissolved by universal rights. Might by itself is never right. It requires qualification through principles of universality. Universal rights ought to have absolute might. This is the question: How can this Ought be turned into an Is? As long as the armed peace of deterrence prolongs the state of war, the equipmentalization of future-oriented human existence, whose freedom, according to Heidegger, resides in its being-possible (Seinkönnen), remains caged in the spirit of militarism, and will continue to determine history with the fatal prospect of bringing it to a halt. Technology intended to open the future is in the process of closing it down.

The 20th century has confirmed Kant’s prophetic foresight. The entry into the nuclear age has brought humanity dangerously close to extinction. The looming threat posed by weapons of mass destruction, the danger of nuclear proliferation defines our age today. Notions of defeat and victory, invoked by political leaders have become meaningless. The destructive power of modern weaponry surpasses the limited nature of the human race both in time (history) and space (earth).

Nietzsche’s Call to Unilateral Pacifism

To many peace advocates Nietzsche seems to be an unlikely ally in their attempt to solve humanity’s predicament. Martin Luther King is not the first or the last to reject ‘the Nietzsches of the world’. What King rejects is not the philosopher Nietzsche, but the National-Socialist misconception of the rejection of violence as cowardice. Yet, precisely this is not what Nietzsche is doing. Nietzsche is not a Social-Darwinist. Key passages in his aphoristic work prove the point. Nietzsche emphasizes the importance of society’s strength to embrace and protect especially those individuals who appear to be ‘weak’ and ‘degenerate’. In their perceived weakness, in their moral and physical handicap, Nietzsche detects the seed for spiritual growth, which alone strengthens a community. The unique perspective so-called ‘degenerate types’ have on life is beneficial to the homogeneous rest, who are blinded by tradition

1‘Beyond Vietnam’, ibid., 161.
The Abolition of War: A Philosophical Imperative

and who resist what Nietzsche calls ‘the plow of evil’, i.e., any change that threatens the tradition. The novelty of ideas, however, will in the long run not endanger race, species or community, but ennoble the human being to a community enriched with the hitherto unknown.

What is true with respect to the individual in its relation to society is equally true with respect to the international community. Nietzsche rejects nationalism and even patriotism as evils that endanger the human race. The contamination of the human mind with ‘mass envy and hatred’ sets the stage for genocide. The task of humanity is the opposite: the dissolution of nations into the strongest possible cosmopolitan mixed race not afraid of any thought and capable of the highest deeds. Of course, Nietzsche here invokes his concept of the overman. The overman can be realized historically by individuals and by a whole people. In each case such realization is a chance of history. It may or may not happen.

The highest and most noble deed is also the most unlikely. The perceived need of military self-defense fuels the arms race by denigrating other nations and peoples as intrinsically evil hypocrites who must by no means be trusted, but be overpowered at the earliest opportunity. In this mindset preemption is the logical extension of deterrence. Just as factual attacks require, as soon as they happen, a self-defensive military response, so the imminent attack merely said to be about to happen calls for anticipatory self-defense.

Conventional wisdom does not lead the way out of this vicious circle. What is needed is courage and wisdom that go against ‘the nature of man’. Here Nietzsche is closer to the thinking of the apostle Paul than he is ready to admit. Paul, too, juxtaposes the ‘natural’ (yucikon) and ‘spiritual’ (pneumatikon) man. The natural man lacks spiritual insight into counterintuitive wisdom, which is why he always falls short of his possibilities and into the traps of violent egoism. Nietzsche and Paul speak so much in one voice that they even use the same rhetoric. The title Nietzsche has chosen for his first aphoristic work *Human, All-Too-Human* (1878) is directly borrowed from the apostle’s writings. When Paul admonishes the community in Corinth not to give in to their ‘all-too-human’ vices he rejects the same kind of addictive greed (pleonexia) as Nietzsche does in his book of that title and elsewhere (*Habsucht, Mehr-und-mehr-haben-wollen*).¹

This is not the only point of agreement between the two. In his recognition of the self-destructive whirl of the vicious circle Nietzsche dismisses militarism as untenable. His vision of peace transcends the whole history of warfare into a possible future of deliberate peace. Such prospect, however, comes at a high price. The call goes to the most militarized nation. It is the radical call to total unilateral disarmament. The future of mankind lies in such a decision. Why the best-armed nation would ever want to engage in such a move is not clear. We may, however, speculate that it is the direct experience of the moral and material cost of militarism that may convince a people of the futility of the whole endeavor. Such recognition can only come as a conversion away from

¹ Cor. 3, 3. 4: este kai kata anqrwpon peripateite; ouk anqrwpoi este; 1 Cor. 15, 42-53; Rom. 1, 29.
old habits into the untrodden realm of a ‘peace of mind’. The conversion comes about in ‘a height of feeling’ that provides courage and ease to think and do the unthinkable. In the light of this conversion the ‘armed peace’ appears as what it is, a contradiction in terms, a ‘truce’ as Kant says, that does not qualify as peace but is a state of war. Built on ‘the absence of peace of mind’, hostilities, as Kant says, unabolished by truce, lock mankind up into the state of nature. The urge to leave such a state of war opens up the only path left to what Nietzsche qualifies as ‘real peace’ and Kant ‘perpetual peace’. To make his point absolutely clear, the Antichristian Nietzsche speaks like a Christian martyr. ‘Rather perish than hate and fear, and twice rather perish than make oneself hated and feared’ is the maxim to be embraced by the most militarized nation first. In his theology of the cross Paul gives the same advice. Both Nietzsche and Paul are aware of the outrageous nature of their suggestion. Self-sacrifice, of which the cross is the Christian symbol, will be rejected by the natural man.¹

Conclusion: Kierkegaard, Heidegger and the Call of Truth

We may agree with Kant and Nietzsche and still find ourselves entangled in humanity’s age-old predicament. We are inclined to ask rhetorically: Which Superpower would voluntarily implement Kant’s articles of perpetual peace? Which commander-in-chief would lay down his arms and destroy the most destructive weapons ‘only’ to give up his hegemonic privileges? The realist Nietzsche is himself torn between pessimism and optimism. Aware of the moral weakness of human nature Nietzsche points at the futility of all disarmament-treaties. All treaties manifest the human tendency to avoid the necessary radical decisions. Humanity must—and perhaps will—be forced to the only solution available—by the catastrophic consequences of its own decadence. The divine spark of humanity will be rekindled when survival depends upon it. ‘Only when this kind of need has become greatest,’ Nietzsche says, ‘will the kind of god be nearest who alone can help here.’ Martin Heidegger, in his Spiegel-Interview from 1966, alludes to this insight by Nietzsche (and Hölderlin) when he declares without a trace of resignation, but with the ‘dis-illusioned’ honesty of modest realism that ‘only a god can save us.’² If anything at all, it is necessity that will bring humanity to its senses. This


²The interview took place in September 1966 and was published after Heidegger’s death in Der Spiegel, May 31, 1976. A translation by Maria P. Alter and John D. Caputo, Philosophy Today
has nothing to do with fatalism. It is a position that strives beyond the
dichotomy of pessimism and optimism.

We know that the ‘kind of god’ the Antichristian Nietzsche is invoking is
the overman who can be both a person and a people. In each case the overman
is a lucky event in history, an accidental incidence, a fortunate coincidence, the
famous kairos that brings its own salvation. How far fetched this possibility
really is Nietzsche makes clear with his choice of metaphors. Only ‘a stroke of
lightning’ can destroy the idolatry of militarism, because only a destruction of
this kind will and can happen ‘all at once’.

In contrast to Nietzsche, Kant does not appear to be more of a realist, but he
is less of a pessimist. His articles of law are pragmatic steps headed towards
peace. Kant does not promise to get there. Their implementation, however, will
give peace the chance it needs to become real in history. Hence, Kant calls
nations to leave the state of nature by submitting themselves to the rule of
international law. Only before an international court of justice can nations
abjure pure force of raw might to be guided by judgments of right. Against
Clausewitz we must conclude, politics has its continuation here, in the power
of law, not in war, which is its failure. War is regress into the lawless state of
nature. As such it calls for continued progress into the peace of the civil state.¹

The imperative of peace is absolute for both Kant and Nietzsche, today more so
than ever. If reason cannot convince humanity, the prospect of annihilation
must do the job. The nuclear age reveals the urgent relevance of both thinkers.
Those who refuse to heed their call reveal their terrorist nature by engaging in
calculations of collateral damage.

Another visionary joins the discussion here. His judgment is
uncompromising, because it is based on a fundamental insight into the human
relation to truth. Heidegger’s thought resonates heavily with him. According to
Heidegger there is no truth beyond existence. To be of any meaning truth must
be located within the human being where the ‘disclosure’ of its revelation into
the ‘unconcealedness’ of allhqeia takes place.² The Danish philosopher and
theologian Søren Kierkegaard was the first to remind us of the intrinsic
connection between truth and existence. In his rejection of Hegel’s concept of
objective truth to be unraveled in a ‘system’ of speculative philosophy,
Kierkegaard posits the ‘subjective truth’ of ‘existence’. However, his
admonition that ‘truth is subjective’ must be understood properly. The

¹This is the call of nature, i.e., nature’s call to humanity to subject itself to the rule of law. In
The Metaphysics of Morals Kant refers to this call of nature as ‘das Rechtsgesetz der Natur’
(‘the law of right intrinsic to nature’, my translation), which Mary Gregor translates as the
der Sitten. Werkausgabe VIII. Frankfurt/Main: Suhrkamp, 374. Cf. § 15 of ‘Private Right’ with
§§ 41, 42 where Kant explicates this call as the ‘postulate of public right’ that proceeds from
private right, and with §§ 43, 44 of ‘Public Right’, where Kant reasons through the necessity of
making the transition from the provisional nature of the natural state to the peremptoriness of
the civil state.

existential fact that ‘truth is subjectivity’ points at the individuality of the subjective thinker. It is the individual subject and subjective individual that contains unique insights into the truths of life and by extension of the world as such. Heidegger captures this idea with his concept of mineness (Jemeinigkeit). In each case existence is mine, that is, a self that speaks from a particular, irreplaceable perspective of history and humanity. The uniqueness of individuality calls for the protection of its absolute truth. Kierkegaard is so adamant in his defense of this subjective truth that he introduces two kinds of realities, the ‘ethical reality’ of the individual subject which can never be known by another, and the ‘conceived reality’ as which anyone is known by another. The discrepancy between these two realities is unbridgeable. The fact that we can never know or understand, i.e. be another human being, calls for absolute respect.

This call is rooted in Genesis 1, 26 according to which humans are created in the image of God. In his seminal work Sein und Zeit (1927) Heidegger makes reference to it. The words contain both the description of a theological fact and the prescription of a philosophical imperative. The fact of godlikeness bears the divine command(ment) to live up to it. Its possibility is real and brings with it an utmost responsibility. Heidegger captures this ‘idea of “transcendence”, that the human being is something that reaches beyond itself’—which literally and philosophically is the meaning of ‘ex-istence’—with his existential-ontological imperative ‘become, what you are!’ Kierkegaard says the same. To ignore this call of humanity is a crime against humanity. To consider any human being dispensable is monstrous. It is immoral to dispense with the particular truths of the individual human nature of truth. Kierkegaard condemns the politician in particular, because the politician is particularly apt and willing to disregard the truth, with the following words: ‘Only a tyrant or an impotent man is content to decimate.’ The judgment addresses modern warfare directly. Modern war depends on the inevitability of so-called ‘collateral damage’. For Kierkegaard such terminology condemns itself. It suggests that the modern age is deliberately ignorant of ‘subjective truth’, and thus prone to declare peace the result of truth annihilation, i.e. mass murder. Peace, however, is, as we have seen, the unconditional imperative of truth. As humans we must heed its call.

Yet modern mankind continues to use ‘war’ and ‘peace’ ‘like two sticks that savages rub together to make fire.’ The ravaging fire of the human history of destruction is still burning. It will continue to do so as long as ‘war’ and ‘peace’ remain defined through each other. Since war is understood as ‘the securing (Sicherung) of peace’, and peace as ‘the elimination (Beseitigung) of war’, mankind is unable to go beyond ‘this war-peace (Kriegsfrieden). Here

---

1Sein und Zeit, 48-49, 145.
Heidegger resonates heavily with Kant, Nietzsche and Kierkegaard. Heidegger condemns ‘war-peace’ as much as Nietzsche the ‘armed peace’ of the ‘absence of peace of mind’ (Unfriede der Gesinnung) and Kant the false peace of truce. This war-peace defines war as a means to peace, as George Bush did explicitly in his 2004 UN address mentioned above. War-peace sees all peace always in need of being secured by war. Peace is just a justification of war, nothing more. War is said to be waged only against war, in defense of peace. Thus, the two depend on each other as each conditions the other into an ever expanding vicious circle of growing violence.

Something is fundamentally out of joint here,’ Heidegger observes. ‘[P]erhaps it has never yet been in joint.’¹ In his Parmenides lecture course some ten years earlier in the winter semester of 1942/43 Heidegger has dug into the problem. With the Romanization of Greek thought the Greek polis of discourse and dialogue was replaced by Roman imperial rule. When military dictatorial command (Befehl) started to determine all domains of life and thought, the thinking of contemplative philosophy lost its forum.² Yet it is here where Heidegger envisions the remedy for mankind. An unapologetic self-assertion of thinking must retrieve the lost heritage of another peace, what Nietzsche calls the ‘peace of mind’ (Friede der Gesinnung). Heidegger finds it in the virtue and comportment of Gelassenheit, which ‘abandons’ all security for the sake of a peace that is ‘released’ to the heart of life and the core of things, to the world as it is. Such peace ‘lets’ world and humans ‘be’ what each is to be in tune with things. This being toward the things themselves is the thinking of being that has been Heidegger’s life-long pursuit. Heidegger’s Gelassenheit is Nietzsche’s ‘height of feeling’ that convinces the worst warmonger of the divine value of insecurity. There is greatness in this courage to be. Its recognition elevates self-sacrifice to the highest maxim. This is the force of the philosophical imperative: to abolish war, possibly at the cost of one’s life, but never of that of another. ‘Rather perish than hate and fear,’ Nietzsche says, ‘and twice rather perish than make oneself hated and feared.’³ The peace of mind of Gelassenheit can come only through the courage of such being.

---

¹Ibid.
³Nietzsche’s aphorism ‘The means to real peace’.
Despite the attention apologies have received by sociologists, linguists, psychologists and many other scientists, philosophers have generally ignored them. In this essay I will first offer the diagnosis that this philosophical inattention is largely due to the belief that apologies are always morally beneficial in view of their therapeutic effects qua remedial strategies. Second, however, I will argue that a closer look at apologies reveals that they have the potential of producing considerable moral harm, and in a variety of cases they do produce such harm. My discussion will be confined to ‘personal apologies,’ apologies that involve a single actor, though with appropriate modification to include group responsibility, what is said here can easily be extended to apologies involving groups.

**Apologies, Accounts, and Dessert**

Apologies are given for a garden variety of reasons in a variety of situations ranging from a mundane minor disturbance to massive harm of whole populations. In all their varieties, nonetheless, apologies seem to present us with the following four fundamental elements:

1) An acknowledgement of a causal association between an incident in question and the apologizer (‘offender,’ ‘actor,’ ‘issuer’)
2) An acknowledgement that the incident in question was bad or inappropriate
3) An admission of responsibility for the harm done by the incident in question
4) An expression of regret for the harm produced by the incident in question

---


2These elements are common to both philosophical accounts in the literature. See, Kort, ‘What Is an Apology?’: G3, and Gill, ‘The Moral Functions of an Apology’:12. Each of these authors adds their own version of a fifth element but this does not affect the purposes of this paper, and
It is generally agreed that apologies are one among four basic categories of remedial strategies where denials, excuses and justifications are the remaining three.\(^1\) Denials function as the first line of defense since they sever the association with the harmful event in question. Justifications and excuses—both known as ‘accounts’—do not deny that a harmful/inappropriate event has taken place, but attempt to demonstrate that there is no blame, no guilt to be assigned to the ‘offender.’ What if, for example, A harmed B only because of an unavoidable accident (excuse) or only because he was defending his country in a battle against B’s unprovoked attack (justification)? Surely, the thought is, A should not be blamed.

Now, as J. L. Austin showed us, it may take protracted philosophical analysis to determine whether we are dealing with an excuse or a justification, and/or whether a given account is successful in alleviating or removing guilt and responsibility.\(^2\) Regardless, the important point is that when a person A provides an account for the occurrence of some harm h, A aims to establish that A is not guilty and no blame can/should be attributed to A for h. And this is the point that makes accounts incompatible with apologies. The giving of an apology by A entails that A has admitted guilt and has conceded that, regardless of whether the outcome of the act in question was bad or not, A is to be blamed for it. It follows that an apology is due only if an account cannot be given, and the other way around: if an account can be given, an apology ought not to be given.

It is interesting that often enough this incompatibility is ignored. For all too often we find people apologizing while at the same time giving an account for their actions. To be sure, this is probably the result of the apologizer attempting to show sympathy and regret for the harm suffered by the victim, but the fact remains that in view of the incompatibility involved, one does not and should not apologize for what one is not morally guilty for. When an account is available, one can (and it would be nice to) empathize but one does not apologize.

This brings me to the question of what it is to have a warranted or deserved apology. This is a complex issue that would take us too far afield. Fortunately, since my aim here is to demonstrate the harms of an apology, it will be sufficient to point out that under certain conditions we have a clearly unwarranted or undeserved apology and issuing this sort of apology results in harm. What are these conditions? I propose that an apology from person A to person B for harm h suffered by B is unwarranted if any of the following obtain:

---

we can ignore it. Similarly, there may be additional elements involved in an apology, but this is not the place for a full blown exploration of the logical territory of the subject.


Apologies and Moral Harm

1) A is not causally, or formally responsible for h
2) An account for h can be provided
3) Harm h is the causal consequent of prior harm h* wrongfully inflicted on A by B.

Condition (1) says that if we cannot trace a causal chain from the harm done to the issuer of the apology and the issuer of the apology is not in a formal relation of responsibility for the acts of the actual offender (one such as, for example, a legal guardian of a child has to the child) then an apology is not warranted. This flows from both elements 1 and 3 of an apology as discussed above, and blocks the mouthpiece syndrome of apologizing for all sorts of things that you have nothing to do with, as for example, apologizing for the deeds of your ancestors, or to foreigners for the idiocy of your countrymen, or even to your house visitor for raining while she visits. Condition (2) flows directly from the incompatibility of apologies and accounts discussed above. Condition (3) needs some explanation.

Suppose that B wrongfully shoots A, and A is now bleeding to death on B’s carpet. Can it be said that B is still entitled to an apology from A for ruining his carpet with his bleeding? Hardly, I believe! In this case, I think, an apology to B is simply unwarranted. B is now suffering the causal consequences of his own initial wrongful action; his carpet is being ruined exactly because he shot A; but for B’s shooting, A’s bleeding and thus the destruction of the carpet would not have occurred. Had B not ‘started it,’ his carpet would not have been ruined and there wouldn’t have been an occasion for an apology. In the context, an apology is not warranted. To put the same point slightly differently: whatever right B may have had for an apology for the harm he is now suffering, he has forfeited it when he wrongfully initiated the chain of events that led to that suffering. Because of his action, he no longer deserves an apology. There is an intuitive reason for this. The aim of an apology is to provide compensation, or reward to the victim of some harm. But in the present circumstances the victim is only the “victim” and in reality the offender. To reward an offender for his wrongdoing makes a mockery of the system of rewards and punishments. And this is exactly what would happen if we offer an apology to B in these circumstances.

Notice that (3) should not be confused with (2). The issue is not that in the case described harm h is excusable or justified and thus an apology is unwarranted, anyway. Rather the point is that perhaps neither an excuse not a justification can be given, and still an apology would be unwarranted if we can trace a causal chain of events that is initiated by one’s wrongful actions and harm ensues as a result of those actions. Consider, for example, the case where country B initiates an unprovoked, wrongful invasion of country A as a result of which the citizens of A take up arms and engage in guerrilla warfare against B. The issue here is not whether the citizens of country A are justified in their reaction. Perhaps the citizens of A engage in measures hugely disproportionate to the harm they suffer from B’s invasion. Say for example they engage in suicide bombings killing hundreds of innocent people at a time, or they manage
to nuke millions of B’s population in retaliation of having lost a few hundred of their own. It is no longer clear that they can provide an account for their actions. What is clear, however, is that in the context of B’s ‘having started the whole thing’ B can hardly complain that A engages in excessive actions not warranted by the rules of warfare and demand an apology for such actions! An apology in this case is plainly unwarranted.

With these qualifications in mind, let us turn to the issue at hand, the harms of apologies.

**The Benefits of an Apology**

At first glance, apologies seem to have all around beneficial effects. Consider the victim first. Every person has a public ‘moral face,’ a record or dossier of sorts that contains the moral image of the person. Ironically, when an offence has occurred the victim’s moral record comes immediately under suspicion. For, after all, it could well be the case that the victim is in some way responsible for the offense in the first place. (A notorious case in point is date-rape where, as a norm, the victim has to defend herself against the suspicion that ‘she went asking for it.’) But it is here that apologies come to the aid of the victim. For the offering of an apology entails admission of guilt/responsibility by the apologizer, and this sets the record straight by ‘clearing up’ any lurking misconceptions that the victim is in some way responsible for what happened. Once the apology is given, the victim’s dossier is freed of any possible spoils. If something harmful happened, it is now shown that it happened to her and this could not in any way diminish her moral image. In addition, the victim gains in personal satisfaction that something inappropriate happened to her and it is now acknowledged as such. The victim has been violated but now moral respect towards her is being restored.

There is an additional benefit to the victim from the offering of an apology. The apology helps heal the moral wounds suffered by the victim and thus paves the way for forgiveness and reconciliation. This is not to say that apologies are necessary for forgiveness and/or reconciliation. A victim may forgive her offender only because she is inclined to do so for reasons of her own (say, she is morally magnanimous, a good Christian) and the decision to forgive is independent of the offender’s apology. And similarly, one may wish to reconcile for pragmatic, circumstantial reasons and do so in the absence of an apology and/or a genuine feeling of forgiveness. Nonetheless, it goes without saying that under normal circumstances (when, for example, one is not too revengeful) an apology surely facilitates the path to healing, forgiveness and reconciliation, all being good and desirable moral outcomes.

When we turn to the offender, again we seem to encounter benefits: It is part of the understanding of an apology that the actor has done something wrong and is now apologizing because of the infliction of that wrong. Having done a wrong, the actor’s public moral image, his ‘moral dossier,’ has been tainted; his behavior made him a lesser member of the moral community in question.
However, by offering an apology, the issuer engages in a remedial, restorative work on his record. For he now attempts to demonstrate that he made a mistake which he regrets and indirectly asks for forgiveness for the harm he has caused. In doing so he shows that he is willing and ready to regain the moral status he had prior to his offense and once again gain acceptance as an equal member of the community. In the absence of a justification or an excuse, the offering of an apology goes a long way towards restoring the apologizer’s moral integrity.

Finally, apologies seem to do a lot of good to the relevant communities involved. The actor’s apology serves as assurance that the actor is sensitive to the harm he has caused, he understands that he has acted outside the standards of the community, and he has thus posed a threat that if left untreated may be repeated in the future. In giving an apology, he demonstrates his moral sensitivity. He shows that with his actions he has caused harm, he makes a gesture of respect for the moral sensitivities of the relevant community, and assures its members that they are safe from similar harm in the future. It is no wonder, as the criminal justice system demonstrates with the practice of apologies in parole hearings, that an apology is an indispensable part for the offender’s readmission to the community.

Thus, apologies seem all-around beneficial, and this could explain the philosophical inattention to the subject. Nonetheless, it must be noted that none of the beneficial effects of an apology need obtain. In the first place, no matter how sincere an apology may be, it is hardly sufficient to change the mind of judges and juries that the victim is also responsible for what happened if the facts, or even the prejudices, speak otherwise. Whether the victim’s moral dossier is in fact restored is hardly determined by the offender’s apology. Similarly an apology is not sufficient for forgiveness. A revengeful enough (or hurt enough) victim can be hardly moved towards forgiveness by the offender’s sincere and deeply felt apology. In the second place, the apologizer may or may not achieve any benefits from apologizing depending on a garden variety of factors, as for example, the public’s perception of genuine remorse involved, the gravity of the offense, the frequency of the offense and/or of the practice of apologizing, and the tolerance of the moral community. Finally, the community benefits from an offender’s apology only if it is given sincerely and not in order to deceive (as it could be the case, for example, when one apologizes to the parole board only to gain sympathy and be set free.) Yet the problem is that the community can never be sure of the sincerity of the apologizer. In the end, the “rosy picture” of benefits may simply not obtain. What is worse, we will now see, apologies in many cases can cause harm and in some cases inevitably result in harm.

The Harms of an Apology

There are apologies and there are apologies and some are given spontaneously by the issuer’s own initiative and some are given after a request. It seems that when we think that apologies are good all around, what we have
in mind is cases that the apologizer issues the apology spontaneously rather
than as a result of a request to provide it, and he does so under apparently
unproblematic conditions: the focus is sharply on the victim, there appears to
be clear and significant harm suffered, and there appears to be no reason to
think that the apology is unwarranted. What I will show now is that even under
this best case scenario, apologies can and do result in harm.

Consider first apologies that are spontaneously, albeit insincerely given in
order to benefit. The criminal justice system provides a case in point. One who
has been found guilty or who is undergoing a parole evaluation has every
reason in the world to appear empathetic, reformed, and overall a better moral
person than his offense otherwise indicates. An apology, no matter how truly
insincere it may be, is almost a panacea solution for the occasion. The problem
that insincere apologies create is that apart from the harm the actor of insincere
apologies produces to oneself (since this adds to his moral demeanor),
significant, additional harm ensues for those who are willing to give him the
benefit of the doubt and believe him. If he is discovered, the victim and the
relevant communities are hurt twice over (once for the original harm from the
offense and twice from being deceived), and if not discovered, wrong prevails
in that those who have been hurt in the first place are now under a state of
deceit.

Second, consider apologies that, even if sincerely given, are given for
reasons of expediency, or by a habitual apologizer, but are not ‘really felt.’
Sometimes an apology is a quick-fix for an otherwise unpleasant but minor
conflict situation which threatens to become more serious than it is warranted
in the absence of a quick apology that puts a stop to the escalation. Then there
are apologies that are the product of habitual apologizing as a quick-fix remedy
for one’s habitual, repeated offenses. Yet an apology ought to be an act of
redemption. As Louis Kort has observed, a necessary condition for an apology
to be genuine is that it involves the making of a gesture of respect towards the
victim as a person that has the right not to be treated in the (offensive) manner
that the actor has treated him.1 To the degree that the actor is not appropriately
sensitive to the gravity of the situation, and he shows little empathy for the
harm he has caused to his victim, to that degree his ‘apology’ does not capture
the appropriate respect for the person’s hurt, and this does nothing other than
add insult to injury for the victim. And it goes without saying that expedient
and/or repeated, perfunctory apologies cheapen the practice and hurt the moral
standards of the community that employs that practice as one remedy for
harmful situations.

When we turn to the issuer of the apology the possibility of harm is even
more pronounced. First, consider the case of an actor that provided a
spontaneous, apparently warranted, deeply felt and sincere apology, and yet,
the apology was not accepted. This could be so for a variety of reasons: it could
be that the victim was too revengeful, or perhaps, too deeply hurt to feel that
the apology is not enough (will not do) for the occasion. Or it can be that the

---

1 Kort, L. ‘What Is an Apology?’: G-11
victim felt that the apology was insincere or not sincere enough. The empathy and the sincerely felt feelings of the actor are in principle inaccessible to the victim and thus are prone to interpretation and possible rejection for reasons other that their veracity. And the trouble is that the actor in those circumstances cannot do anything about the rejection. Yet by contrast, she is liable to additional harm resulting from the rejection. For it is more than humiliation that is involved here; it is also the public perception of either the additional gravity of the situation as shown by the victim’s degree of hurt (a degree high enough that turned down the apology) or the alleged insincerity of the actor, or both. In any case, the actor gets hurt, and she can do nothing about it.

Second, spontaneous, apparently warranted apologies that are hastily given can easily result in harm for the apologizer. If the apology is requested, at least there is one more person of the belief that an apology is due for the occasion. By contrast, when the decision is based entirely on the apologizer’s own judgment, personal psychology may lead the way. It could be that the genuine regret the apologizer feels for what has happened obscures her judgment or otherwise forces her to a quick apology expressed primarily as a gesture of empathy for the suffering of the victim; and that can be true even if on cooler assessment of the situation an account of the harm could be given (or perhaps is given alongside with the apology, as for example tends to be the practice with apologies accompanied by excuses). But this can be a grave mistake. One surely harms oneself (at least) when one admits guilt where guilt is not due.

Third, harmful results can occur for the apologizer if the expression of the apology goes a long way beyond what is called for in the occasion. An apology is a remedy. As such, there ought to be an appropriate degree of fit between the seriousness of the moral harm caused by the offense and the expression of apology offered in return for that harm. Yet there is always a risk that the apologizer becomes too apologetic; that is, he ‘overdoes it’ by exhibiting largely incommensurate amounts of regret and/or offering excessive compensation and/or promises for future restoration. These may be all due to some subjective feeling of guilt, yet the objective harmful effect is there, namely the public perception of the offense as much greater than what in fact is and thus the greater the stain on the moral face of the offender.

Finally, some harm for the issuer of an apology is always inevitable. This is because an apology demonstrates once and for all the admission of guilt and that, in turn, produces a permanent stain in the issuer’s record. Each person’s reputation, his moral face as it were, is dynamically altered by every action he undertakes. The admission of guilt that is an essential element of an apology forever alters that record, and there is nothing one can do to make it disappear in the future. This stigma is now a fact that will haunt the future. And we should not forget that we have created a monster of sorts given that now we have provided a legal ground for the victim to request punishment and/or compensation for the harm that, to the apologizer’s admission, he has suffered.¹

¹ To be sure, it could be replied that the actor deserves the moral stigma that his admission of guilt produces, since this is simply just retribution for the harm that his offense created in the first place; and similarly, with the legal repercussion, if any. Be it as it may, the trouble with
All in all, even under the best circumstances apologies seem to hide more harm that is at first apparent. When we turn to requested apologies things become even graver.

Requests for Apologies

Requested apologies can be particularly harmful. When a request for an apology is issued, someone other than the apologizer has made the judgment that an apology is due. Indeed, for the issuer of the request for an apology, the matter is closed: the addressee has been tried and found guilty, an apology has been determined to be warranted in this case, and all that remains is the call to the addressee to endorse these findings. This is a tall order for the addressee, who is already ‘cornered’ into a fundamentally defensive position from the very issuance of a request. To be sure, there are situations when an apology is obviously warranted. But it is equally sure that there are many situations where this sort of determination is a very complex affair requiring some degree of philosophical sophistication in order to be gotten right. Yet every request for an apology implies either that the situation is patently obvious or that the required sophistication is there. This is hard to swallow, at least with regard to the average issuer of a request for an apology in many cases. Yet the addressee of the request is called to succumb to this judgment. Moreover, in the run of the mill case, the addressee cannot easily take her time to assess the appropriateness of the request. Once the request is issued the clock is ticking so to speak. The presumption of guilt is already built into the fact that a request for an apology has been issued. The longer one waits the more the risk of strengthening the impression that she is guilty after all. And all this pressure is, of course, magnified if it comes on top of the dictates of one’s own conscience or psychology that may already be placing its own share. When an apology is requested, the specter of coercion looms large.

The relative positions of the issuer of the request and the addressee can be decisive. Even when the issuer of the request for an apology is in some sense in an inferior position vs. the addressee (e.g., an ordinary citizen vs. an authority, a child asking a parent, a subordinate asking the boss, etc), the pressures from the request as discussed above may well be thought of as coercive. But when the request is issued from superiors to inferiors, the coercion is blatant. Regardless of the ‘or else’ clause being explicitly pronounced, the fact is that this “request” is really a command backed by threats and coercion is implicitly operative. For what are the alternatives available to the apologizer? To opt for declining the apology is to opt for consequences more severe than the humiliation and the loss of moral face that the admission of guilt implies. If the superior is the boss, the actor may also lose the job; indeed he could perhaps be grateful that only an apology is being requested since it could have been the lesser punishment if punishment was intended. Similarly, if the offender is a
Apologies and Moral Harm

child and the request comes from a parent, teacher, or other superior, it may well be that the apology is the least evil. If the offender is a criminal under judgment by the parole board, it can be almost guaranteed that the parole will be denied if he is unapologetic.\(^1\) In every case, the harm is there, for coercion is coercion no matter the benefits. To the degree that an apology is not the product of one’s free will, under any conception of justice, is to suffer great moral harm.

What if the request for an apology is made public, or even worse, is made by the public? Clearly the possibility of succumbing to pressure and apologizing under coercion increases multifold, since the collective power of the community to coerce knows no limits. And there is an additional, albeit major danger in this case: the public is rarely in possession of the fuller description of the events. But in the absence of this information how is it possible for the public to determine correctly if a requested apology is unwarranted when in fact it may not be? And even if the requisite information is there, how plausible is it to assume that the public has the requisite sophistication to assess correctly complex cases of unwarranted apologies? The addressee of a public request for an apology, it seems, is always in grave danger of finding herself in the unfortunate position of being under undue pressure to apologize when rightly she should not.

Conclusion

There is a nightmare scenario, I believe, where a request for an apology results in necessary harm: Suppose we are confronted with a public demand for an apology from a superior, yet the apology is not warranted because the harm suffered by the requester is the causal consequent of his own prior wrongful infliction of harm to the addressee (situation 3 earlier). Here the real offender is the one that turns the tables, as it were, and issues a de facto command for an apology he does not deserve. What is one to do? To resist giving an apology, unwarranted and undeserved as it may be, still creates the impression of a hard-core offender who resists recognizing the harm he inflicts. So, in view of the built-in presumption of guilt present in the already issued request, one ends up being morally stigmatized twice over! On the other hand, the overall moral damage seems greater if one yields and apologizes though the apology is not deserved. First, the apologizer himself suffers damage. Given that the apology is not deserved, if the actor gives in to pressure and apologizes for pragmatic reasons, the actor loses respect for himself because of his absence of moral courage. Second, to repay the wrong that is done to him with the reward of apologizing is to obliterate the value of rewards for doing right and being rewarded for doing right. The system of rewards and punishments is damaged,

\(^1\)It can hardly be said that this is not a relevant case because the criminal justice system does not explicitly issue a request for an apology, or that the whole system of sentencing and of parole does not expect an apology in order to show leniency. And it can hardly be doubted that the level of pressure to be found here does not amount to coercion.
and any morally sensitive person should feel bad about that. Third, it seems that in yielding to the real offender’s pressure, he rewards the offender and he may thus reinforce bad behavior and indirectly ask for more! So, here is the dilemma: decline to apologize and suffer the harmful personal consequences, or cause harm to all by apologizing? I do not know whether we should say that, since declining to apologize causes less overall harm, the morally right thing is to bite the bullet and confess guilt, or whether we should say that the innocent should not be made the sacrificial lamb for the greater good. These are large questions. But at least that much should be clear: one way or the other, in the present predicament moral harm is simply unavoidable. All in all, apologies are not that morally innocent after all.
Blackburn’s Supervenience Argument against Moral Realism

Makoto Suzuki

To examine Blackburn’s argument\(^1\), we need to understand what notion of supervenience he employs. Blackburn’s supervenience thesis is stated in terms of deflationary truths or facts (Blackburn, 1993, 130). I take Blackburn to characterize ‘truth’ or ‘fact’ as follows: a judgment that P is true [or states a fact] if and only if P, where ‘P’ is (a) a sentence in which a general term is applied to a constant or bound variable or to an ordered set of them, or (b) a sentence somehow constructed of a set of such sentences. Blackburn thinks that by characterizing the notion of fact or truth in this way, realists and anti-realists can agree on the same supervenience thesis; realists can read (a) above as an ascription of a real property to an object or objects while anti-realists can avoid that realist rendering. Anti-realists can claim that the general term in (a) refers just to such a non-substantive property that for anything x, x has the property of being D if and only if Dx.\(^2\)\(^3\)

Let me use Blackburn’s supervenience theses in his *Spreading the Word*.\(^4\) Blackburn endorses the following supervenience thesis:

\[(S1) \quad \Box ((\exists x) (F_x \land G^*x) \to (y) (G^*y \to F_y))\]

‘\(\Box\)’ is necessity operator. ‘\(F\)’, ‘\(G\)’ and ‘\(G^*\)’ are predicates. ‘\(F\)’ stands for a moral property (for instance, being evil) and ‘\(G\)’ for a natural property (for instance, being a killer), and \(G^*\) is ‘some definite total set of G truths.’ (Blackburn, 1993, 131) In ‘Supervenience Revisited’, Blackburn also formulates another supervenience thesis, which he does not endorse. If we revise it in a way parallel to (S1), we get:

\[(?1) \quad \Box ((\exists x) (F_x \land G^*x) \to \Box (y) (G^*y \to F_y))\]

\(^1\)The fullest exposition of Blackburn’s argument is given in ‘Supervenience Revisited’ (Blackburn, 1993). However, Blackburn also gives the argument in other places.

\(^2\)This is the case of monadic property.

\(^3\)There is a controversy over whether projectivists like Blackburn can coherently hold a deflationary theory of truth, fact and property. See, for example, Horwich (1993) and Smith (1994). However, this issue is not crucial for our discussion.

\(^4\)The formulations of supervenience in ‘Supervenience Revisited’ are problematic (Drier, 1992, 17).
The difference between (S1) and (?1) lies in the point that (?1) has an additional ‘□’ before (y) (G*y → Fy). That is: (S1) says within each respective possible world, if x and y are naturally alike, they are morally alike; and (?1) says if x and y are naturally alike, they are morally alike across every possible world – they cannot morally differ. (S1) permits, for example, that in this world any act of torture is wrong, but in another possible world no such act is wrong; in fact, every such act is right in this other world. This scenario is excluded by (?1).1

As far as I understand Blackburn, he takes ‘F’ to stand for any moral property, so he should have bound F by a universal quantifier. As for ‘G*’, Blackburn talks of ‘the sense of ‘G*’’, and states that ‘G*’ is ‘a complete, base description of a thing, …, telling everything that could be relevant to determining its A-state.’ (Blackburn, 1993, 133) ‘A-state’ is what we here express by ‘F’-fact (Blackburn, 1984, 183). Such a collection of natural properties is a G*, and the properties in a G* might differ in each particular thing that has F, as Blackburn says, ‘There is no claim that G* provides the only way in which things can become F’ (Blackburn 1993, 131). Thus, ‘G*’ is another variable, which Blackburn should have bounded with an existential quantifier; moral supervenience means that for every moral property, there is some set of natural properties that underlies it.2 Then, (S1) should be reformulated as follows:

(S2) \( \Box (F \text{ in } M) \left( \exists G^* \text{ in } N \right) \left( (\exists x) (Fx \& G^*x) \rightarrow (y) (G^*y \rightarrow Fy) \right) \),

where M is a family of moral properties, N is a family of natural properties, and (F) and (\( \exists G^* \)) are the universal quantifier for F and the existential quantifier for G* respectively. Accordingly, (?1) is rewritten as:

(?2) \( \Box (F \text{ in } M) \left( \exists G^* \text{ in } N \right) \left( (\exists x) (Fx \& G^*x) \rightarrow \Box (y) (G^*y \rightarrow Fy) \right) \)

‘\( \Box \)’ expresses analytical or ‘competent’ necessity. Blackburn claims the supervenience of moral facts is a conceptual matter (Blackburn, 1993, 137). However, he thinks even if Quine is right and there is no distinction between analytic and synthetic truths, supervenience is at least a matter of ‘competence’ (ibid., 142-144). Both ‘analytic’ necessity and ‘competent’ necessity need some explanation, but he only gives it for the latter. Blackburn takes competence to be ‘half conceptual but perhaps half scientific’; ‘competently possible worlds’ are ‘those that are as a competent person might describe a world as being’ (ibid., 142). Then, presumably, ‘competently necessarily P’,

1For a useful discussion on the implications of different versions of supervenience, see Kim (1993) and McLaughlin (1995).
2If we use the universal quantifier instead, we will have the claim that if something satisfies a moral description and a natural description, everything that satisfies the natural description satisfies the moral description. This would implausibly exclude the coincidental co-satisfaction of a moral and a natural description.
where \( P \) is a statement, means that there is no world that a person competent both in language and in science can take to be possible and where \( P \) is false.

Now, if nothing is \( G^* \) and \( F \), \( \exists x \) \( (Fx \& G^*x) \) is false, and so both (\( F \) in \( M \)) \( (\exists G^* \text{ in } N) \) \( ((\exists x) \ (Fx \& G^*x) \rightarrow (y) \ (G^*y \rightarrow Fy)) \) and (\( F \) in \( M \)) \( (\exists G^* \text{ in } N) \) \( ((\exists x) \ (Fx \& G^*x) \rightarrow (\Box y) \ (G^*y \rightarrow Fy)) \) are vacuously true. Thus, the necessitation of each, that is, (\( S2 \)) and (\( ?2 \)), will be vacuously true if nothing could be \( G^* \) and \( F \); then, it seems that ‘being virtuous supervened upon being homogeneously made of granite’ just because nothing could be homogeneously made of granite and be virtuous. This is absurd. Thus, Blackburn confines the scope of \( F \) and \( G^* \) to the cases where it is possible that something is both \( G^* \) and \( F \) (Blackburn, 1993, 132-133).

**Blackburn’s Explicit Supervenience Argument**

Blackburn’s explicit argument consists of two steps: A. argument for (\( S2 \)), and argument for the analytical possibility (\( P \)), namely, that something is \( G^* \) but not \( F \); and B. argument that it is mysterious that both (\( S2 \)) and (\( P \)), so it needs some explanation for why we are to accept them, but that realism cannot provide such an explanation. Let us look at them in order.

**2-A. Blackburn’s Argument for the Combination of (\( S2 \)) and (\( P \))**

Actually, Blackburn does not defend (\( S2 \)) among possible formulations of the supervenience of moral truths.\(^1\) What Blackburn provides instead are reasons for the claim a. that supervenience, where the necessity involved is analytic or competent, is to be accepted, and for the claim b. that we may not want to endorse (\( ?2 \)). I am willing to grant that a., that is, some thesis of the supervenience of moral facts on non-evaluative facts is an analytic or at least competent necessary truth for this reason or others. So I will only examine Blackburn’s argument for b. that we may not want to endorse another supervenience thesis (\( ?2 \)).

Blackburn has us suppose (\( ?2 \)). Now, as we saw in section one, there is a reason to confine the scope of \( F \) and \( G^* \) to the cases where it is possible that something is both \( G^* \) and \( F \). Then, for every \( F \), there is a possible world where (\( \exists x \) \( (Fx \& G^*x) \)). So, together with (\( ?2 \)): \( \Box (F \text{ in } M) \ (\exists G^* \text{ in } N) \ ((\exists x) \ (Fx \& G^*x) \rightarrow (\Box y) \ (G^*y \rightarrow Fy)) \), we will have \( \Box (y) \ (G^*y \rightarrow Fy) \) for every \( F \). Thus, Blackburn says, accepting (\( ?2 \)) is ‘in effect nothing but a roundabout way of committing ourselves to (\( N \))’, where (\( N \)) is (\( F \text{ in } M \)) \( (\exists G^* \text{ in } N) \) \( (\Box x) \ (G^*x \rightarrow Fx) \) (Blackburn, 1993, 134)\(^2\). It gives us the reduction of supervening facts to natural facts, which, first, supervenience claims are supposed by many not to

---

\(^1\)I think Kim (1993) gives better formulations of supervenience. However, this is not crucial to Blackburn’s argument.

\(^2\)I reformulate modal claims in ‘Supervenience Revisited’ in order to remove the problems pointed out by Drier (1992, 17).
give\(^1\), and second, is ‘markedly less attractive’ in the moral case (ibid., 134). The third point is simply that we do not accept (N), if we, as Blackburn does, construe the necessity operator as expressing analytic necessity (ibid., 137-138).\(^2\)

All the three points are controversial. As for the first, many philosophers think moral supervenience gives reduction.\(^3\) They think supervenience gives that kind of reduction which allows the multiple realization of supervenient properties, and (N) is just such a reduction thesis: ‘(N): \((F \text{ in } M) (\exists G^* \text{ in } N) \Box (x) (G^*x \rightarrow Fx)\)’ does not entail ‘(F in M) \((\exists G^* \text{ in } N) \Box (x) (Fx \rightarrow G^*x)\)’ Since our intuitions are divided on this issue, they give no reason to avoid (?2) or (N) rather than (S2).

As for the second point, Blackburn says nothing specific, so let me treat it as the different expression of the third point. Blackburn says in rejecting (N):

\[
\text{For it is not plausible to maintain that the adoption of some particular standard is ‘constitutive of competence’ as a moralist. People can moralize in obedience to the conceptual constraints that govern all moralizing, although they adopt different standards and come to different verdicts in the light of a complete set of natural facts. (Blackburn, 1993, 137)}
\]

Blackburn seems to forget how he has characterized ‘G*’ or ‘F’. According to him, ‘G*’ is a complete natural description of each thing, telling everything that could be relevant to determining its moral status (Blackburn, 1984, 183). ‘F’ is characterized to stand for any moral property. Suppose that (N): for every F, there is some G* such that \(\Box (y) (G^*y \rightarrow Fy)\), and that it is (analytically) possible that something is G* and F. Even then, what at most follows is that it is analytic that for every moral property, there is some set of natural properties that necessitates the moral property; or that every moral description is entailed by some natural description. The formula does not say what particular thing satisfies the description of ‘G*’; it neither says what particular moral property F is. People do adopt different standards and come to different verdicts in the

\(^1\)On the other hand, as Blackburn says, (S2): \(\Box (F \text{ in } M) (\exists G^* \text{ in } N) ((\exists x) (Fx \& G^*x) \rightarrow (y) (G^*y \rightarrow Fy))\) does not give reduction even given that for every F, there is a possible world where (\(\exists x\) (Fx \& G^*x)); for something that is G* could be F.

\(^2\) As Zangwill (1995) points out at 257, if the inner necessity operator of (?2): \(\Box (F \text{ in } M) (\exists G^* \text{ in } N) ((\exists x) (Fx \& G^*x) \rightarrow (y) (G^*y \rightarrow Fy))\) expresses metaphysical and not conceptual necessity, (N): \(F \text{ in } M) (\exists G^* \text{ in } N) \Box (x) (G^*x \rightarrow Fx)\) does not express a conceptual truth. Blackburn’s arguments against (N), as you will see, can affect only the conceptual version of (N). I agree with Zangwill that the above version of (?2) is plausible (Zangwill, 1995, 249 & 257-258; see also Kim (1993), 59-61). However, this is not crucial to Blackburn’s argument, for the argument does not rely on the denial of (?2), where the inner necessity operator expresses metaphysical necessity. This version of (?2) is as acceptable as (S2) for Blackburn’s purpose, because, given that it is conceptually (or competently) possible that (\(\exists x\) (Fx \& G^*x)), neither denies (P): it is conceptually (or competently) possible that something is G* but not F.

\(^3\)For example, see Jackson (1998), 122-125.
Blackburn’s Supervenience Argument against Moral Realism

light of natural facts, but it is consistent with the above supposition; they disagree with what thing satisfies ‘G∗’, and with what moral property F is.

Blackburn thinks some version of the open question argument successfully refutes the claim that every description of moral fact is entailed by some description of moral facts (Blackburn, 1998, 317). Roughly speaking, the open question argument goes like this. For any claim that G is conceptually identical to or entails F, where F is a evaluative description and G is a natural description, we can sensibly ask, ‘Is G really F?’ Therefore, every such claim is false.1 The open question argument is controversial2 and disputed by the naturalistic realists who argue that moral terms can be defined by naturalistic terms.3 Because this argument is what Blackburn ultimately relies on, his rejection of (N) and hence of (?2) is not securely grounded.4

The open question argument is simultaneously crucial to Blackburn’s argument for (P), namely, it is conceptually (or competently) possible that something is G∗ but not F. The argument has been suggested above where Blackburn argues against the claim (N). The argument is that because some version of the open question argument repudiates any entailment from natural facts to a moral fact, so we get (P), namely, it is conceptually (or competently) possible that for some F and G*, something is G* but not F. Thus, the success of the open question argument is crucial for Blackburn’s argument for the combination of (S2) and (P).

2-B. Blackburn’s Puzzle(s)

Suppose, for the sake of argument, both (S2): □ (F in M) (∃G* in N) ((∃x) (Fx & G*x) → (y) (G*y → Fy)) and (P), and consider why Blackburn thinks this combination is mysterious (Blackburn, 1993, 134-135). The explicitly stated version of Blackburn’s puzzle can be appreciated if you tentatively imagine the necessity and possibility operators in (S2) and (P) express not analytical or competent modalities, but other modalities, such as metaphysical ones. Now, the G*/F worlds – the worlds where anything that is G* is F – and the G*/O worlds – the worlds where anything that is G* is not F – are permitted. However, due to (S2), the G*/F or O worlds – the worlds where

1This argument originates from Moore (1902), Ch. 1.
2See Miller, 2003, 10-25. Note that if the necessity operators in (?2) are taken to express not analytical necessity but competent necessity, it is even unclear whether the open question argument is relevant to the version of (?2).
3See, for example, Jackson (1998), 151-152.
4If we concede that some version of the open question argument works, and hence it is false that any moral description is entailed by some natural description; then, by reductio, we end up denying that (N); for every F, there is G* such that □ (y) (G*y → Fy). Thus, given that it is reasonable to confine the scope of F and G* to the cases where it is (analytically) possible that something is G* and F, we should also reject (?2): □ (F in M) (∃G* in N) ((∃x) (Fx & G*x) → □ (y) (G*y → Fy)). Thus, if we find attractive the conceptual or competent version of the supervenience thesis in moral area, and if we ignore the formulations of supervenience by other authors than Blackburn, we have no option but to accept (S2). Note that, however, this argument crucially depends on a controversial claim that the open question argument refutes the claim that every moral description is entailed by some natural description.
something is G* and F while something is G* but not F – are not permitted. We naturally wonder why. Surely one answer is that such exclusion is a brute fact, but this is ad hoc and leaves us wondering whether we have any reason to believe in such a fact. Without some explanation and hence reason at hand, we had better withhold the acceptance of the exclusion of the G*/F or O worlds; if so, we should put either (S2) or (P) in abeyance.¹ Now, return to (S2) and (P), where modal operators represent analytical or competent modality. Is the combination of (S2) and (P) still puzzling? Blackburn answers affirmatively (Blackburn, 1984, 137).

But we should answer negatively. If we have already had (S2) and (P) as conceptual or competent truths, there is a reason to accept the combination; the meaning of moral language (together with our scientific knowledge, in the case of competently necessary truths) gives us a reason to accept both (S2) and (P). Unlike metaphysical or physical necessities or possibilities, conceptual or competent necessities or possibilities entails the availability of explanation for why we are to accept them. The reason we are to accept the necessity of the proposition that a bachelor is an unmarried adult male is given by the meaning of the words it contains,² or by the fact that the competent speakers of English agree that it is impossible to be false. The reason for accepting the necessity of the proposition that water is H₂O is given by linguistically and scientifically competent people’s convergent beliefs (about chemistry). Since the modality of both (S2) and (P) are conceptual or competent, we have reasons to accept them by definition. Thus, accepting the combination of (S2) and (P) is unproblematic.

Let me put my point in another way. Blackburn’s puzzle goes as follows. It is possible (p) that something has certain natural properties and has a moral property. It is possible (q) that something has the same natural properties and does not have the moral property. However, given moral supervenience, it is possible (p & q) that something has certain natural properties and has a moral property, and something else has the same natural properties and does not have the moral property. Why is that? This is not a real puzzle if the involved modalities are conceptual or competent. It is possible (p) that Makoto is a bachelor. It is possible (q) that Makoto is married. It is impossible (p & q) that Makoto is a bachelor and married. This trio of modal statements is not mysterious. For they are all analytic truths, which are explained in terms of meaning. It is possible (p) that the liquid I am having is water. It is possible (q) that the liquid I am having is not H₂O. It is impossible (p & q) that the liquid I am having is water and not H₂O. This trio of modal statements is not mysterious. For they are all competent truths, which are explained in terms of

¹Suppose we reject (S2) and instead accept that version of (?) in which the inner necessity operator expresses not conceptual (or competent) necessity but metaphysical necessity. Then, Blackburn’s puzzle can be restated as follows: why is it conceptually (or competently) impossible that something is G* and F while something is G*and not F, though it is conceptually (or competently) possible that something is G* and F and it is conceptually (or conceptually) possible – though metaphysically impossible – that something is G* and not F?

²This point is also made by Jackson (1998), 125.
Blackburn’s Supervenience Argument against Moral Realism

linguistically and scientifically competent people’s convergent beliefs about chemistry.

However, perhaps Blackburn’s real puzzle exists elsewhere. Blackburn (1993, 137) suggests the real problem of realism is that only projectivism can accommodate the fact that the role of moralizing is to be choosing, commending or forbidding on the basis of natural properties; and so realism fails to provide a proper reason to take (S2) as a (normative) constraint of moralizing. Blackburn’s demand is to explain how realists can accept the right kind of reason for (S) coherently with their theoretical framework.

Look first at Blackburn’s claim about projectivism. He seems to say that projectivism takes the whole role of moralizing to be guiding desires and decisions on the basis of natural properties, so that the role is analytically or competently embedded in our moral language; and that therefore, supervenience thesis (S2), analytically or competently interpreted, is accommodated within the projectivist framework (Blackburn, 1993, 137). There is disagreement about whether the whole role of moralizing is to guide the desire and choices: it might also consist in describing moral facts. However, it is plausible that guiding the desires and choices on the basis of natural properties is part of the role of moralizing, and that the fact explains the sensibility of something like (S2).¹

Are realists unable to accommodate that reason for (S2)? Is there any trouble for realists to admit that part of the role of moralizing is to guide desires and decisions on the basis of natural properties?² If, as Blackburn thinks, they have to say that the whole role of moral thinking and discussion is to describe moral aspects of the world, then realists are committed to denying our intuition that part of the role of moralizing is to guide the preferences and decisions. However, realists need not say the whole role of moralizing is description.

If moral beliefs satisfy the following two conditions, then moral thinking and discussion can guide desires and decisions on the basis of natural properties. 1. Motivation Condition: if a healthy person has a moral belief, then (at least in many cases) she tends to feel, desire, and act in a certain definite direction. For instance, if one believes that killing is wrong, she (on many occasions) tends to feel angry with killers, to feel guilty when she kills somebody, and to desire to withhold from killing and to stop others from killing. 2. Natural Ground Condition: a healthy person occasionally changes her moral belief on the basis of natural properties and upon her reflection and discussion with others. For example, if a smoker comes to be convinced that smoking in public space has a certain natural property (say, harming others), she has some tendency to change her moral belief that smoking in public space

¹Zangwill argues that it is ‘utterly unhelpful’ to refer to the point of moralizing in explaining moral supervenience as a conceptual truth (Zangwill, 1995, 246). However, if what is required is to explain its normative aspect, then explaining it by reference to the point of moralizing might well be helpful.

²Non-naturalistic realists might have a reason not to admit this point. But then I am unsure whether these realists should anyway embrace the supervenience of the moral on the natural.
is morally permissible. The satisfaction of Motivation Condition ensures that moral judgments guide desires and decisions; and the further satisfaction of Natural Ground condition ensures that it is done through moral thinking and discussion on the basis of natural properties. The formulations of the two conditions need further clarification, but they work for my purpose. If moral realists can provide some account of moral beliefs that satisfies these two conditions, why cannot moral realists admit part of the role of moralizing is to guide desires and decisions on the basis of natural properties? I would never deny that it is a challenge for moral realists to discern and defend such a theory of moral belief. However, at least Blackburn provides no argument that it is impossible.

Let me recapitulate. First in 2-A, we see Blackburn’s argument for (S2) and (P), where the modal operators are analytical or competent. We find that his argument relies on the open question argument, which is controversial and begs the question against some realists. Conceding (S2) and (P) for the sake of argument, we next see Blackburn’s argument which is supposed to show that accepting the combination of (S2) and (P) needs explaining, and that projectivists can provide one but realists cannot. However, it turns out that if (S2) and (P) are conceptual truths as Blackburn asserts, we have explanations for (S2) and (P). Still, Blackburn’s real point seems to be that there is an intuitive explanation for the sensibility of the moral supervenience constraint, and that while projectivism can incorporate this explanation into its theory, moral realism cannot do so. However, if realism provides a plausible theory of moral beliefs that fulfills two conditions, namely, Motivation Condition and Natural Ground Condition, it will accommodate the intuitive explanation. Thus, so far Blackburn’s argument is inconclusive.

Argument from Moral Truths and the Representational Nature of Belief

In his early essay ‘Moral Realism’, however, Blackburn suggests a reason why realists cannot incorporate the intuition for the conceptual versions of the supervenience thesis. Blackburn again contrasts the ability of projectivists to explain why we are to accept the supervenience thesis with realists’ inability to do so (Blackburn 1993, 113-115). While conative attitudes are not supposed to represent truths or facts in the inflationary sense, beliefs are; and while projectivists take the mental states of moral judgments as some conative attitudes, realists take them as beliefs. Blackburn thus argues that projectivists need to explain only why we are to accept the supervenience of the content of a conative attitude that something is F, where F is a moral concept, upon a set of natural properties. However, realists must explain the supervenience of a moral truth that something is F upon a set of natural properties, as well as the supervenience of the content of a belief that something is F upon the set of natural properties. Now, there is a good reason for the claim that analytically, the content of a moral judgment that something is F supervenes on a certain set of the natural properties. Our practice requires the consistency of the contents
of moral judgments over the things of the same natural properties, which distinguishes them from mere caprices. For projectivists, that is sufficient for explaining what they are supposed to account for, namely, the supervenience of the content of a moral judgment (taken as a conative attitude) upon a set of certain natural properties. However, for realists, that is not sufficient because they have to explain the supervenience of a moral fact on a certain set of the natural properties, and the above reason does not support that. Moreover, if realists have no reason to expect that a moral fact supervenes on a set of the natural properties, because the contents of moral beliefs are supposed to represent moral facts, they have to admit that the contents of moral beliefs might not supervene on a set of the natural facts.

Realists must accept and explain both the supervenience of the contents of moral beliefs and the supervenience of moral facts in some non-deflationary sense. They must admit them because, as Blackburn points out, there is a good reason to believe in the supervenience of the contents of moral judgments on natural facts, and if they are beliefs, then unless moral facts really supervene on natural facts, we are systematically mistaken. Suppose that we make a moral judgment that a particular person deserves punishment on the basis of a certain natural property of S’s, e.g., his offending people by ruthlessly revealing their secrets. Further suppose this judgment is true. Now, suppose we make another moral judgment that another person deserves punishment on the same ground. If a moral property supervenes on a set of natural properties, this judgment might well be true, too, since the natural property we take as the ground might well underlie the moral property that someone should be punished. However, if not, there is no guarantee for that. Indeed, since we use general rules, such as that killing a human being in normal situations is wrong, if a moral property (say, wrongness) does not supervene on a certain set of natural properties (say, killing a human being in normal situation), then we might well be systematically and massively mistaken. I am unsure if such an error can be corrected unless we have a special ability of intuition to detect moral properties directly, without reflection on natural properties of the things judged. However, such intuitive ability is mysterious.1 These are unbearable consequences, so realists must admit the supervenience of moral truths on a set of natural properties.

Let me relate this argument with the result in 2-B. At its end, we tentatively concluded that moral realists might be able to coherently accommodate an intuitive reason for supervenience thesis, namely, that part of the role of moralizing is to guide our desires and decisions on the basis of natural properties. However, this intuition does not support the supervenience of moral truths on a set of natural facts. If so, the argument in this section says, because realists have found no reason to believe that supervenience, and because moral beliefs are representational, the supervenience of the contents of moral beliefs on natural properties is also suspicious.

---

1On the other hand, suppose that we somehow correct such an error; if it is analytic that a moral fact supervenes on the natural properties, in making such a correction we are committing a conceptual mistake (Sigrún S. Svavarsdóttir points this out).
Then, realists have to explain why we are to believe the supervenience of moral facts on a certain set of the natural properties. If realists just assert that the supervenience is a brute fact, we have no warrant to believe them. We should consider what explanation realists can offer.

Explanations from Moral Realists

There is one merit for accepting the supervenience of moral facts on a set of natural facts. It explains and vindicates the following fact: people occasionally try to identify what natural properties make moral truths, for example, what natural properties make it the case that copying and pirating software is wrong. If moral truths did not supervene on a set of natural facts, this investigative activity would not make sense.

This account is not entirely reassuring because our inquiry is perhaps misguided. Realists have another explanation for the supervenience of moral facts. Blackburn agrees that our moral statements supervene on a set of natural properties. Realists can use this datum in two ways to establish the supervenience of moral facts upon a set of natural properties.

The first approach relies on a certain conceptual claim. Realists can claim that any moral term ‘F’ conceptually implies that F supervenes on a set of natural properties. Provided that meaning constrains the kind of property referred to by ‘F’, it is analytically guaranteed that F supervenes on a set of natural properties if there really is such a supervening property. In this case, the supervenience of moral facts or properties on a set of natural things is both analytic and substantive, provided that there are such moral facts and properties. This approach burdens realists to empirically repudiate the possibility of error theory, that there is no property that corresponds to a moral term ‘F’. Possible counter-evidence against this threat might be that the application condition of each moral term can be given roughly in natural terms.

The second approach relies on inference to the best explanation. Realists may say the datum that moral judgments supervene on sets of natural facts is best explained by the fact that the moral facts really supervene on those sets or at least the similar sets, and that we have roughly tracked the combinations of the supervening moral facts and those sets of natural properties in moral thinking; through the history of this tracking, we have acquired the moral

---

1Klagge (1988, 467-469) holds that even if the supervenience of moral facts is a brute fact, we may cling to it. His reasoning runs as follows. The judgments of mental properties analytically supervene on behavioral properties in the same way the judgments of moral properties analytically supervene on natural properties. It is justified to believe that mental facts metaphysically supervene on physical properties and that it is a brute fact. Thus, by analogy, it is justified to believe that moral facts metaphysically supervene on natural properties and that it is a brute fact. However, it is false that the judgments of mental properties analytically supervene on behavioral properties. It makes sense to ascribe different mental states to those individuals who behave in the same way in response to the same input. For the inner processes that produce the behavioral output in response to the input might differ between different individuals.
perspective that if someone issues different verdicts across the things with the same relevant natural properties, then we do not take these verdicts as moral judgments. This is why it has become necessary that moral judgments supervene on natural facts. The success of this approach depends on whether moral supervenience is discovered and transmitted.

You can combine the two approaches. However, to take the second discovery-and-transmission approach and claim that it is conceptually (rather than competently) necessary that moral judgments supervene on natural facts, you must concede one thing: until the supervenience became a conceptual truth, our ancestors had made no moral judgment (in the current sense of the phrase).

In any way, moral realists have something to prove. However, they have ways to explain the supervenience of moral truths (and hence of moral beliefs) on natural properties, and Blackburn has not shown that these explanations cannot work.¹

**Bibliography**


¹I thank Sigrún Svavarsdóttir for helpful comments.