Editorial Board

Carol Nicholson
Rider University
USA
Donald Poochigian
University of North Dakota
USA
T. Ann Scholl
United Arab Emirates University
UAE

Board of Reviewers

Chrysoula Gitsoulis
City College, CUNY
USA
Keith Green
East Tennessee State University
USA
Effie Papoutsis Kritikos
Northeastern Illinois University
USA
Raymond Petridis-Tzomanos
The New School for Social Research
USA
Penelope Voutsina
University of Sheffield
UK
AN ANTHOLOGY OF PHILOSOPHICAL STUDIES

VOLUME II

Edited by

Patricia Hanna
# Table of Contents

**List of Contributors**

**History of Philosophy**

**Introduction**

Hanna, P.

**Part I. Ancient and Medieval**

1. Aquinas and Ibn Rushd on the Added Dimension of Reason in the Comprehension of Revelation in Two Medieval Religious Tradition
   Ahmad, Isham P.

2. The Conquests of Alexander and Greek Knowledge of Indian Philosophy: A Context of Understanding
   Bruseker, George

3. Vasubandhu's use of Vijñapti-mātra: A Reinterpretation in the Context of the Bodhisattva Ideal
   Collinson, Hazel

4. The Hollow Truth of Ockham's Razor
   Goodman, Jeffrey

5. Socrates' Wisdom in the Apology
   Green, Jonas Stigaard

6. The Aristotelian Method and Aristotelian Metaphysics
   Tahko, Tuomas E.

**Part II. Modern**

7. Kant's Rejection of Right of Revolution
   Azeri, Siyaves

8. Spinoza on Hatred and the Impotence of Reason
   Green, Keith

9. Kant's Analytic-Synthetic Distinction: A Reply to a Classical Objection
   Lee, Seung Kee

10. Class-theoretic Paradoxes and the Neo-Kantian discarding of Intuition
    Onof, Christian

11. Kantian Intelligibility
    Räsänen, Petri

12. Throwing Out Scholastic Metaphysics with the Bathwater?
    Scholl, T. Ann
### Part III. 20th Century

13. Wittgenstein's Treatment of Philosophical Problems  
   Gitsoulis, Chrysoula  
   Page 137

   Igrek, Apple Zefelius  
   Page 149

15. Levinas and Arendt on Violence and Conscience: The Disharmony of Everyday Life  
   Kaye, David  
   Page 159

16. Describing the Unspeakable and Demonstrating the Unprovable: Wittgenstein, Gödel, and the Metamathematics that Wasn't  
   O'Loughlin, Ian  
   Page 171

### Part IV. Metaphysics, Epistemology and Philosophy of Language

17. Internal Realism  
   Bremer, Manuel  
   Page 181

18. The Fregean Theory of Belief Reports and Anaphora  
   Bonardi, Paolo  
   Page 191

19. How Private is Private?: What's Wrong with Chomsky's Mentalism  
   Hanna, Patricia  
   Page 199

20. Lotteries, Knowledge and Vagueness  
   Murzi, Julien  
   Page 211

21. Abstraction and the Interpretation of Quality  
   Poochigian, Donald V.  
   Page 223

22. Hegel's Analysis of Universality in The Science of Logic  
   Trisokkas, Ioannis D.  
   Page 235

### Part V. Ethics and Political Philosophy

   Heilinger, Jan-Christoph  
   Page 249

24. Should we believe our Ethical Theories? Virtue Ethics and Self-Effacement  
   Martinez, Joel A.  
   Page 261

25. The Physician-Patient Relationship: An Aretaic Formulation  
   McCabe, Matthew S.  
   Page 271

26. Utilitarianism, Obligation, and Significant Sacrifice  
   Michael, Mark A.  
   Page 281
<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Author</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>28.</td>
<td>Competition and Reasons of Autonomy: A Rejection of Korsgaard’s Theory of Reasons</td>
<td>Salazar, Heather</td>
<td>303</td>
</tr>
<tr>
<td>29.</td>
<td>'Weeds are the Bane of the Fields; Desire is the Bane of Mankind': Desire, Intentional Action, and the Good Life</td>
<td>Schiltz, Elizabeth A.</td>
<td>313</td>
</tr>
<tr>
<td>30.</td>
<td>What are the Prospects for Armchair Ethics?</td>
<td>Winter, Michael</td>
<td>327</td>
</tr>
<tr>
<td></td>
<td><strong>Part VI. Philosophy of Religion</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32.</td>
<td>The Question of Ineffability in Semantics and Negative Theology</td>
<td>Lehtonen, Tommi</td>
<td>349</td>
</tr>
<tr>
<td>33.</td>
<td>Deconstructing God-On the Logical and Conceptual Clarification of the Problem of Evil</td>
<td>Muenster, Harald</td>
<td>359</td>
</tr>
<tr>
<td>34.</td>
<td>An Evaluation of Freud’s Concepts of Morality and God</td>
<td>O’Meara, William</td>
<td>371</td>
</tr>
</tbody>
</table>
List of Contributors

Isham P. Ahmad is an Assistant Professor of Philosophy at the International Islamic University, Malaysia. His research interest is mainly in epistemology. His primary focus concerns the relationship between revelation and reason, and how this shapes our worldview and our understanding of ethics. He has applied this research to interfaith dialogue, collaborating in works on the understanding and appreciation of religion from a global perspective.

Siyaves Azeri is a Research Fellow in Philosophy at the Iranian Institute of Philosophy, Tehran, Iran. He received his Master of Arts in philosophy from the New School for Social Research, and his PhD from the University of Ottawa. His major areas of research concern the problem of personal identity and the metaphysics of the self, and the application of particular philosophical theories to literature and film.

Paolo Bonardi is a PhD student in the Philosophy Department at the Université de Genève, Geneva, Switzerland. He has been a visiting student and tutor at the University of Sheffield and New York University. He has delivered papers on epistemology, problems of belief and philosophy of language at a number of philosophical conferences; in addition he has published papers papers in Dialectica and Abstracta.

Manuel Bremer is University Lecturer in Philosophy at the University of Düsseldorf. He has also worked as a software engineer and computer science teacher. He studied Philosophy, Linguistics and Computer Science at the universities of Cologne, Hagen and Düsseldorf included philosophy, linguistics and computer science, receiving Master of Arts, a PhD and a post-doctoral thesis [German "Habilitation"]. He specializes in logic, philosophy of language and cognitive science. He has published on semantics, rationality, information flow, logic, cognitive science, and epistemology.

George Bruseker is a PhD student in Philosophy at the University of Athens, Greece. His research interests include ancient Greek, Indian and Chinese philosophy and the theory and practice of intercultural philosophy. He is currently completing a dissertation on the encounter of Greek and Indian philosophy at the end of the classic and beginning of the Hellenistic period. He has a BS in history and philosophy from the University of Alberta, Canada, and an MA in philosophy from the Catholic University of Leuven, Holland.

Hazel Collinson is a PhD student in Philosophy at the University of Manchester, UK. Her research focuses on issues of ‘consciousness’ in the Sanskrit philosophical traditions of Vasubandhu (fourth century CE), Kumārila (seventh century CE) and Śaṅkara (eighth century CE). Other research interests include epistemology, philosophy of language and pedagogical methods in Sanskrit philosophical texts. She has a BA and an MA in the study of Religion and Theology from the University of Manchester.

Chrysovali Gitsoulis received her PhD in Philosophy from the Graduate Center of the City University of New York in 2008. Her research interests include Wittgenstein, ancient Greek philosophy, and contemporary moral and social issues. She holds a BA in Philosophy, a BS in Mathematics, and a Minor in Classics from Rutgers College, New Jersey. She is an adjunct lecturer at the
City University of New York, and is currently working on publishing her dissertation.

**Aleksandr Golozubov** received his PhD in Philosophy of Culture and Philosophical Anthropology from V.N. Karazin Kharkiv National University, Ukraine, where he is an Associate Professor of Philosophy. He has been a visiting scholar at the Central European University, Budapest, the Institute for Advanced Studies in the Humanities, University of Edinburgh, the Catholic University of America, Washington, D.C., as a Fulbright Scholar. He has published papers on Theology of Laughter in the western culture; Franciscan tradition, the problem of the sacral violence and utopian and anti-utopian genres.

**Jeffrey Goodman** is an Assistant Professor of Philosophy at James Madison University, Harrisonburg, Virginia, USA. He works in Metaphysics, Philosophy of Language, Philosophy of Mind, Logic, Epistemology, and Philosophy of Religion, and has published papers in these areas in international journals.

**Jonas Stigaard Green** is a Sessional Lecturer in Philosophy at Birkbeck College, London, UK. He received his PhD in Philosophy from University of London at King’s College. His research interests include Ancient Greek Philosophy, Virtue Ethics and Moral Psychology, and Selfhood and Character.

**Keith Green** is an Assistant Professor of Philosophy at East Tennessee State University, Johnson City, Tennessee, USA. He serves as the Advisor for Religious Studies. His work is in Philosophy of Religion, ethics and religion, philosophical issues related to hate and hatred, and Western religious thought.

**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 Chomskyan theoretical linguistics; she is also developing an realistic ontology of interpretation which finds it roots in the late writings of Wittgenstein.

**Jan-Christoph Heilinger** is a Researcher in Philosophy at the Berlin-Brandenburg Academy of Sciences and Humanities. He is currently a visiting Research Fellow at the Institute for Biomedical Ethics at the University of Zurich. He received his MA from Humboldt-University Berlin. His research interests include ethics, bioethics, philosophical anthropology and philosophy of mind.

**Apple Zefelius Igrek** is a Lecturer in Philosophy at Central Washington University, Ellensburg, Washington, USA. His interests range from Georges Bataille and Michel Foucault to pragmatism and the philosophy of religion. His most recent article is ‘Concrete Containment in Late Capitalism, Mysticism, the Marquis de Sade, and Phenomenological Anthropology’.

**David Kaye** is a PhD student in Philosophy at the New School for Social Research, and an Adjunct Professor of Philosophy at St John's University, New York City, NY, USA. He received his BA in Philosophy from Towson
University, and his MA in Philosophy from the New School for Social Research. He has published on Nietzsche, Heidegger, Levinas, Husserl and Arendt. His current research focuses on the role of animality in the history of metaphysics and post-metaphysics.

Seung-Kee Lee is an Associate Professor of Philosophy at Drew University, Madison, NJ, USA. His Ph.D. degree is from The Catholic University of America. His research interests include Kant’s theory of judgment, on which he has published several papers, and the logical and epistemological theory of Kant and the German idealists. He is currently working on the problem of the synthetic a priori in Kant and German idealism. He is a contributor to the forthcoming Kant-Lexikon.

Tommi Lehtonen is University Lecturer in Philosophy in the Faculty of Public Administration at the University of Vaasa, Vaasa, Finland, and Adjunct Professor of Philosophy of Religion at the University of Helsinki, Helsinki, Finland. His areas of special interest include philosophy of religion, philosophy of language, and philosophy of science. His current research focuses on the nature and functions of religious language, philosophical questions related to secularization and the privatization of religion, and philosophy of comparative research. He is the author of Punishment, Atonement and Merit in Modern Philosophy of Religion.

Joel A. Martinez is Assistant Professor of Philosophy at Lewis & Clark College in Portland, OR, USA. He works at the intersection of Ethical Theory, Psychology and the History of Ethics. His research focuses on the practical importance of moral theory; specifically, he is interested in the relationship between moral theory and moral education. He has recently written and spoken on such topics as the development of Stoic ethics, theories of moral learning, as well as morality and the problem of demandingness. Currently, he is working on a series of papers on virtue ethics.

Matthew S. McCabe is Assistant Professor of Philosophy and Religion at Washington College, Chesterton, Maryland, USA. His fields of research are theoretical and applied ethics. He is currently working on virtue ethics and its application to issues in professional ethics.

Mark A. Michael is Professor of Philosophy in the Department of History and Philosophy at Austin Peay State University, Clarksville, Tennessee, USA. His research interests are in environmental ethics, ethical theory, and political theory. He has edited an anthology on various approaches to saving wildlife entitled Preserving Wildlife, and published work environmental ethics, ethical theory, and political theory.

Harald Muenster is a PhD student at the University of Munich, Munich, Germany. In addition he studies German and English as preparation for teaching in secondary schools. He studied philosophy at Rutgers University, and received his MA from the Munich School of Philosophy in Munich based on a thesis on the mind-body problem. He does research in the philosophy of Johann Gottlieb Fichte.

Julien Murzi is a PhD student in Philosophy at the University of Sheffield, UK. He has an MA in Philosophy from the University of Rome “La Sapienza”. He has studied at St Louis University (USA), and St Andrews University (UK). His main areas of interests are logical revision, intuitionism, the realism/anti-
realism debate, and the semantics of future contingents discourse. He is currently working on a thesis on intuitionistic revision. He has published papers on revisionary arguments and paradoxes of knowability.

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 pragmatism and education.

Ian O’Loughlin is a PhD candidate at the University of Idaho, Moscow, Idaho, USA. He received his BA in Philosophy, University of Minnesota, Morris. He is a Research Assistant in the Interdisciplinary Epistemology Project in the Department of Philosophy at the University of Idaho, and has been a Kennedy Scholar in the Department of Classics. He holds an appointment as Philosophy Instructor at North Idaho College, College, Coeur d’Alene, Idaho. He works in ethics, critical thinking and epistemology.

William M. O’Meara is Professor of Philosophy at James Madison University, Harrisonburg, Virginia, USA. He received his PhD in Philosophy from Loyola University. His research interests include process philosophy, American Pragmatism, phenomenology, and Marxism. He has published work on critical thinking for high school students.

Chris Onof is Associate Research Fellow and teaches philosophy at Birkbeck College, London, UK. He received his PhD in Philosophy from University College London. He is also Senior Lecturer in Engineering Mathematics at Imperial College London. His philosophical interests lie chiefly in Kant’s transcendental idealism, Kantian ethics, the 'hard' problem of consciousness, and existential phenomenology. He is joint founding editor of Episteme, a new journal of Social Epistemology.

Donald V. Poochigan 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.”

Petri Räsänen is a Researcher in Philosophy in the Faculty of History and Philosophy at the University of Tampere, Finland. He received his PhD in philosophy from the University of Tampere (2005) and MBA from the University of Vaasa, Finland (1993). His present research interests include Immanuel Kant’s philosophy and political philosophy (including topics such as freedom, liberalism, and economic-political matters).

Heather Salazar is Assistant Professor of Philosophy at Spring Hill College, Mobile, Alabama, USA. She received her PhD and MA in Philosophy from the University of California, Santa Barbara; she holds a BA in Philosophy and a BS in Psychology from Arizona State University. She works in Meta-ethics, Normative Ethics, and Business Ethics, and has published on Kantian Ethics.
Elizabeth Schiltz is an Assistant Professor of Philosophy at the College of Wooster, Wooster, Ohio, USA. Her research focuses on Ancient Greek Philosophy, Nietzsche and East-West Comparative Philosophy.

T. Ann Scholl is an Associate Professor of Philosophy and Chair of the Department of Philosophy at the United Arab Emirates University, Al-Ain, UAE. She has held positions at the American University of Kuwait, Kuwait City, Kuwait, Minnesota State University, and Moorhead University, USA. Her research interests include Early Modern Philosophy, feminism, and social issues in Islam. She has published on Early Modern Philosophy, feminism, epistemology, and is the author of Descartes' Dreams: Imagination in the Meditations.

Ioannis Trisokkas is a Research Fellow in the Department of Philosophy at the University of Warwick, Warwick, UK. He has a BA in Philosophy and Psychology from the University of Ioannina (Greece) and an MA in Continental Philosophy from the University of Warwick. He also studied Philosophy of Language and Mathematics at King’s College London (with a University of London scholarship) and Speculative Logical Theory at the University of Tübingen (Germany). He works on Kant, Hegel, and Ancient Greek Philosophy, as well as metaphysics, epistemology, and philosophy of language.

Tuomas E. Tahko received his PhD in Philosophy from Durham University, UK. His research is on metaphysical realism, contemporary analytic metaphysics, modality, and essentialism. He has published on a priori knowledge.

Michael Winter is Associate Professor of Philosophy at University of St. Thomas St. Paul, Minnesota, USA. He does research on ethical theory, and is currently working on a book about a new approach to virtue ethics.
Introduction

The present volume is a collection of papers selected from those presented at the 2nd International Conference on Philosophy sponsored by the Athens Institute for Research and Education (ATINER), held in Athens, Greece 4—6 June 2007.

Held annually, this conference provides a singular opportunity for philosophers from all over the world to get together and share ideas with the aim of expanding our understanding of our discipline. At the 2007 conference, philosophers from 23 countries (Australia, Austria, Canada, Colombia, Denmark, Egypt, Finland, France, Germany, Greece, Iran, Italy, Kuwait, Malaysia, New Zealand, Russia, Sweden, Taiwan, Turkey, Ukraine, United Arab Emirates, United Kingdom and the USA) met and talked with one another in the shadow of the Acropolis.

Over the course of the conference, 103 papers were presented. The 34 papers in this volume were selected for inclusion after a process of review by at least two of the editors and reviewers. 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 gives some sense of the variety of topics addressed at the conference. However, it would be impossible in an edited volume to ensure coverage of the full extent of diversity of the subject matter and approaches brought to the conference itself by the participants, some of whom could not travel to one another's home countries without enormous difficulty.

Patricia Hanna
Editor
USA
Part I

Ancient and Medieval
Aquinas and Ibn Rushd on the Added Dimension of Reason in the Comprehension of Revelation in Two Medieval Religious Tradition

Isham P. Ahmad, International Islamic University, Malaysia

Medieval philosophers were confident that they must and could interpret and assimilate philosophy—specifically Greek philosophy—into the religious ethos. The medieval philosophers' belief that it is necessary to incorporate philosophy into the religious ethos is that philosophy provides proofs for theoretical opinions in religion. Al-Farabi argues that "[t]heoretical opinions in religion have their proofs in theoretical philosophy, while they are taken in religion without (argumentative) proof.”¹ In order to reconcile philosophy with a religion that is revelation-centric, philosophers had to find a way to incorporate revelation into philosophy. Their attempt to do this resulted into one of medieval philosophy's major contributions to the advancement of philosophy.

Why were they interested in the relevance of philosophy to religion and what are the benefits from this arduous endeavor? Were Plato and Aristotle so intellectually attractive that the medievals found them irresistible? Or did they see a far greater benefit in to this endeavor, bordering on an almost dire need not to only to reconcile, but to reunite, revelation with reason? Is such a synthesis possible? Why is it desirable? How, does this synthesis elevate the presentation and understanding of the religious ethos? Questions which weighed heavily on them still have repercussions today. This synthesis has implications for the ideas of natural law's impact on human relationship structures through its redefinition of the concept of justice and the resultant

¹al-Fārābī, Kitāb al-millah wa nūfūl ukhrī, ed. M. Mahdi, (Beirut, 1968), 47.
change in our understanding of our personal religious beliefs and the concept of religion itself.

The medieval desire to interpret and assimilate philosophy into their religious ethos may not only benefit one party but that philosophy itself could benefit from this endeavor, if the integration is successful. The Muslim philosophers following Aristotle\(^1\) argued that for philosophy to be truly beneficial, it must not only benefit the few, the small number of philosophers who studied it, but it must benefit the many, society at large. For example, Al-Farabi considered that religion not only to provide intellectual philosophical answers for the masses but elevated them in accordance with their intellectual capacity, thus providing the potential for continuous elevation to higher intellectual levels. Thus, this endeavor to interpret and assimilate philosophy into their religious ethos is crucial for both reason and revelation.

The Added Dimension of Reason in Ibn Rushd

Ibn Rushd\(^2\) concluded that all knowledge is the knowledge of God and thus, he sought to elucidate this relationship and expound its practical implications. He sought to establish the unity of truth and fought his battle to clear the name of philosophy by struggling to reconcile reason to revelation. By the time of Ibn Rushd, the venture of philosophy had been brought into ill repute and had been discredited by it battles with theologians, culminating with Al-Ghazali\(^3\).

Ibn Rushd struggled to reestablish the importance of philosophy. His counter-offensive occurred on three fronts: legal, theological and philosophical. This counter-offensive was aimed at establishing the legitimacy of philosophical endeavors in the eyes of the religious, but, more importantly, he wanted to established the usefulness of philosophy, along with its contribution

\(^1\)Aristotle says, "For even if the good is the same for the individual and the state, the good for the state is clearly the greater and more perfect thing to attain and to safeguard. The attainment of the good for one man alone is, to be sure, a source of satisfaction; yet to secure it for a nation and for states is a nobler and more divine." Joachim, H. H. Aristotle: Nicomachean Ethics, (Oxford: The Clarendon Press, 1952), 1094b. See also, Ernest Baker, The Political thought of Plato and Aristotle (New York, 1958), 284-292.

\(^2\)Kiki Kennedy Day says that “Ibn Rushd, better known in the West as Averroes, is considered not nearly as influential in the Islamic world as he was in medieval Europe. Here, either because he lived on the Western periphery of the Islamic world or because he wrote such extensive commentaries on Aristotle, he became renowned Latin translations of Ibn Rushd's texts were available in Europe within a century of his death. Coming from a family of eminent jurists, Ibn Rushd had legal as well as philosophical training”. Aristotelianism in Islamic philosophy, [http://www.muslimphilosophy.com/ip/rep/H002](http://www.muslimphilosophy.com/ip/rep/H002)

\(^3\)Oliver Leaman concludes that “although al-Ghazali is often regarded as the archenemy of philosophy, it is evident on closer inspection of many of his texts that he himself seems to adhere to many of the leading principles of Ibn Sina's thought. Also, in common with many other opponents of philosophy, he had a high regard for logic (which was regarded as a tool of philosophy rather than a part of it) and insisted on the application of logic to organized thought about religion. Islam, concept of philosophy in, [http://www.muslimphilosophy.com/ip/rep/H006](http://www.muslimphilosophy.com/ip/rep/H006)
Aquinas and Ibn Rushd on the Added Dimension of Reason in the Comprehension of Revelation in Two Medieval Religious Tradition

to religious understanding. His attempt to do this, raises, and answers, the question of why we should incorporate philosophy in a religious discourse.

Ibn Rushd’s defense against the vilification of philosophy by the great theologian Al-Ghazali comes down to two issues: 1) the Quran itself, which commands the study of philosophy, and 2) the view that when there is an apparent conflict between truths of scripture and the demonstrative truths of philosophy, the literal scripture should be allegorically interpreted to remove the conflict.

I will only provide an outline of Ibn Rushd legal defense here because it has been more thoroughly covered elsewhere. What has been neglected in the analysis of Ibn Rushd’s defense is the world view we can extrapolate from his defense of philosophy. This defense answers the question of whether philosophy is permitted by religion, as well as providing a glimpse into the contribution of philosophy. I argue that if we examine these answers more closely, we can extrapolate the basic presuppositions that Ibn Rushd held and go some way toward reconstructing his world view. Even if it is not a more productive way to explain why Ibn Rushd insisted that philosophy is an important discipline in Islam, I believe that, at the very least, it provides another way to look at seeking answers to this of reconciliation. Therefore, while in the course of discussing Ibn Rushd’s response to the legal indictment against philosophy, I will look for the underlying assumptions and try to draw out from this as much as possible to discover whether we can obtain a better comprehension of Ibn Rushd's views of how the world works, thus, to construct his worldview.

Both revelation and reason claim to provide knowledge. The articles of faith is a form of knowledge of something, what one ought to believe as true. But the sources of knowledge of revelation and reason are different. However, this does not necessarily mean that knowledge from revelation should only be conceived as an emotional attitude or essentially non-intellectual. Ibn Rushd and the other medieval philosophers certainly did not envisage revelation as a non-intellectual affair. It could be easily be argued that most theologians would agree with this conclusion.

The basic foundation for Ibn Rushd's philosophy, like Aquinas, is that the universe is intelligible. This means that the universe has been created with structure and laws that can be comprehended by the finite mind of man. However, since the human mind is finite, its understanding can never be equal to the comprehension of God, the infinite. Nevertheless, human understanding is sufficient for man to understand his own cause and his ultimate destiny. Tranoy concludes that the medieval philosophers had “a faith, however modest, in the powers of the human intellect to understand man and the universe”.

Ibn Rushd argued that it is the Quran itself that commands the study of philosophy. Numerous Quranic verses command man to use his intellect to understand and appreciate God and His creation. Quranic verses such as “Reflect, you have a vision” (59.2) and “they give thought to the creation of

2Tranoy, 100.
3Tranoy, 99.
heaven and earth” (3:191), command man to reflect upon God and His creation. These illustrate that the Quran sees value in the intellectual endeavor. The Quran recognizes that if one were to reflect on the creation, one would confirm the truths of the creator, in other words, the truths of the revelation, showing the Quran itself argues against the non-intellectual

Ibn Rushd argues that the Quran commands the study of philosophy since he equates philosophy with intellectual reflection. Ibn Rushd defines the study of philosophy as the study of existing beings and reflection on them as indications of the Artisan, and since God is the ultimate Artisan, it is a reflection on God and His creation. Therefore it is not only not harmful, but should be either obligatory or at least recommended by the Law. For Ibn Rushd like Aquinas concludes that all knowledge is the knowledge of God, either directly through the study of God, or indirectly through the study of His creation which inevitable reflect its creator, God. This counters the accusations of Al-Ghazali that the study of philosophy is harmful for the Muslims.

Ibn Rushd concludes that there is more than one route to God; philosophy is simply another route to God. Not only is philosophy one of the routes to God but it is a precursor to any understanding of religion as a concept. Therefore, philosophy, Ibn Rushd argues is a precursor to the proper understanding of revelation itself. Far from being harmful, the Shariah, Ibn Rushd argues, encourages and renders obligatory upon individuals who are capable, the study of philosophy. Since all religions require man to transcend the sensory world and find meaning from this transcendency, it is only when one accepts that there is more to life than the life of the flesh and senses can religion begin.

Since it has now been established that the Law has rendered obligatory the study of beings by the intellect, and reflection on them, and since reflection is nothing more than inference and drawing out of the unknown from the known, and since this is reasoning or at any rate done by reasoning, therefore we are under an obligation to carry on our study of beings by intellectual reasoning.2

Ibn Rushd concludes that since the Law has rendered obligatory the study of beings by the intellect, and reflection on them, this implies use of rational inferences (qiyaṣ ’aqli). Thus, the Law establishes the legitimacy of rational speculation (naẓar), whose method reaches perfection with demonstrative syllogism (burḥān). Since, therefore, such obligation exists in religion, then a person who has the capacity of “natural intelligence” and “religious integrity”

---

Aquinas and Ibn Rushd on the Added Dimension of Reason in the Comprehension of Revelation in Two Medieval Religious Tradition

must begin to study philosophy. While most ordinary members of society are convinced through persuasion of the obligations of religion, it is only the philosopher who truly understand the implications of the demands of revelation in other words, it is the philosopher who can truly comprehend and interpret the demands of revelation. Thus, the Quran prefers that we should have demonstrative knowledge. The question remains why is it the Quran made this preference and simply blind faith as the ideal.

Ibn Rushd’s argument against the conflict of revelation and reason is based on the foundation that truths cannot conflict. Truths are truths and, therefore, could never conflict, whatever their source. Since the nature of philosophy is the search for truth, and revelation is the ultimate truth, Ibn Rushd, concludes that the truths of philosophy cannot conflict with scripture (i.e. Qur'an). If the truths of scripture could not conflict with the demonstrative truths of philosophy, all or any such conflict between the two, must therefore be only apparent.

What should be done when these apparent conflict appears? Since allegorical interpretation of scripture is common among the lawyers and theologians and has been long accepted by all Muslims, there is no reason why philosophers also should not resort to this accepted device. Thus, when there is an apparent conflict between truths of scripture and the demonstrative truths of philosophy, the literal scripture should be allegorically interpreted. This is what theologians had done. For example, theologians have always interpreted the verse “The Jews say: 'God's hand is tied up.' Be their hands tied up and be they accursed for the (blasphemy) they utter. Nay, both His hands are widely outstretched: He giveth and spendeth (of His bounty) as He pleaseth.” (5: 67), not to mean that God has hands and therefore a limited physical form. They resort to an allegorical interpretation where the hands of God interpreted as an expression of the power of God. Thus, if the theologians can reinterpret literal Quranic verses when they are apparently in conflict with theological truths and if this is acceptable to the majority of Muslims, there should be no objections to the reinterpretations carried out by philosophers.

If, when philosophy and scripture disagree, scripture should be interpreted allegorically, does it follow that philosophy has priority over scripture? This may lead one to misunderstand and, wrongfully, conclude that Ibn Rushd argued for the priority of philosophical truths over scriptural truths. Theologians had reinterpreted the scriptures allegorically whenever the literal meaning conflicted with the theological understanding. However, the theologians did not think that this entailed that they believed that theological truths had priority over scriptural truths. No one ever objected to the theologians practice of reinterpreting scriptural text. Thus, it would seem hypocritical to accuse the philosophers of believing that philosophical truths were superior to scriptural truths when they engage in the practice of allegorical interpretation accepted by the theologians. All the philosophers hope to achieve by this is what the theologian aimed for, that is to demonstrate the unity of truths: that no truths, whatever their source, can conflict.

The question therefore is not whether the scriptural text could be allegorically reinterpreted, but rather what, how and on what basis this
reinterpretation is to be carried out. This is crux of the problem. For Ibn Rushd, the scope of allegorical interpretation is not open to all subjects but only in areas where there is no unanimous agreement on the meaning of the scripture. If there is no consensus about a particular passage, then its meaning is open for interpretation. Establishing unanimous agreement is much more difficult than most assume.

Philosophers, Ibn Rushd concludes, are endowed with a unique method of learning: that of acquiring beliefs through demonstrative arguments and then them with allegorical interpretation. This method adds new dimensions to the understanding of revelation. It not only confirms the beliefs in revelation, thus serving to confirm the revelation, it also adds to the richness of the revelation by comprehending it in greater depth.

The Added Dimension of Reason in St. Thomas Aquinas

Both al-Farabi and Aquinas grappled with the relationship of their faith to their philosophy in their theory of knowledge. The basic question is this: what is the relationship between the insights revealed to man by the grace of God and the insights man struggled to obtain for himself through his own efforts, i.e. reason? Aquinas’s theory of knowledge differed from al-Farabi’s. Al-Farabi tried to integrate Plato’s theory of knowledge with Neoplatonic thought, with Aristotle influence in the background, placing all three within the Islamic tradition. Al-Farabi produced a complex a theory of knowledge which regarded revelation as a progression stemming from reason. Thus, revelation, was viewed as an extension of reason-an expression of reason.

Aquinas, by contrast, accepted the Aristotelian concept of human knowledge; therefore, his struggle with the role of revelation was more difficult. Aristotle did not consider revelation because he did not have a concept of prophecy. Thus, the question Aquinas had to deal with was the relationship between the Aristotelian concept of human knowledge and the knowledge brought by the Grace of God: viz., revelation. Underlying Aquinas' thought is, of course, the medieval assumption that truth is one and, therefore, that the truths obtained either through reason or revelation could not contradict nor conflict with each other.

The question Aquinas first had to contend with concerned the scope of the two sources of knowledge, reason and revelation. Do they both address the same issues or not? If they do, does one become redundant or complimentary? The next question concerns the method utilized by these two sources of knowledge? The two methods of obtaining knowledge is certainly not the same. Because they utilized used different methods of obtaining knowledge, do they have different ways of reaching the same conclusion like seeing two sides of the same coin, thus the only difference is being the means? Aquinas saw the

2Ibid. 15. ibid. 52.
Aquinas and Ibn Rushd on the Added Dimension of Reason in the Comprehension of Revelation in Two Medieval Religious Tradition

need and more importantly, the value of this attempt to integrate faith with reason. Tranoy pointed out three trends of thought on the issue with Aquinas adding a fourth one. The first, faith alone is sufficient, in fact reason may be harmful to one’s faith.

Three different ways of dealing with the relationship between faith and reason, theology and philosophy, will be distinguished here as a background for an understanding of the fourth, that of Aquinas. In the second century, Tertulian’s attitude to the problem has been epitomized in the phrase “Credo quia absurdum” (“I believe because it is absurd”, though this is not his exact wording). In his teaching the revealed insights of Christianity made any kind of knowledge superfluous. “With our faith we desire no further belief. For, this is our palmary faith, that there is nothing which we ought to believe besides”.

This position is extreme but it is very attractive to many and one should never underestimate its influence on believers. This view is attractive because the faith revealed by God is complete and needs no additional effort, thus its simplicity provides comfort and certainty.

A second view gives priority of faith over reason but sees limited value to reason because the revelation from God is received by man and understood through his reason.

A second, and probably a more important tradition which also gives priority to faith over reason, is connected to Augustine. This tradition recognizes the need for certain rational endeavors and also admits that knowledge through reason is possible, given certain conditions. These conditions lie in the Christian faith itself and in the divine grace and assistance accorded to the believers.... Man’s “natural light”-his intellect or reason-must first be lit by God; then and only then, can man use this faculty to throw light on that which he desires to understand. Therefore, faith in the dogmas of religion must come first; the credibility of Christianity is not dependent on rational proof (which nevertheless this tradition produced in abundance), the proofs being intended to show that the dogmas are logically necessary in themselves. Why, then should intellectual understanding be desirable when it is not necessary for faith? This question ask for the rationale of the credo ut intelligam tradition. Revelation – the Scripture – is the word of God, and words are addressed to and properly received by the understanding. Now, the complete and perfect understanding of the word of God as the Truth can, at best, be attained in the state of beatitude after death.

Here, reason is conceived as the means to understand faith. It is, however, limited by the limitations of human understanding and human comprehension of the revelation. Man will fully comprehend the truth of the revelation only at

---

1For an elaboration of the Tertulian primacy of Faith position, refer to Etienne Gilson, Reason and Revelation in the Middle Ages, New York, 1938. See Chapter 1, “Primacy of Faith”.
3Ibid, 100.
the state of beatitude after death. What is the need to for human further efforts of human speculation and reflection? Is it at best superfluous? Augustine answers no, but you might be surprised by his reason. Augustine held that Neo-Platonism had served him as a preparation for the Gospel by liberating him from his previous materialism; throughout his life he continued to use the modes of thought he learned from this Neo-Platonism.¹

The third trend comes from the Muslim philosophers. This is so because Aristotle was initially transmitted to Christian Europe via the Muslim philosophers’ interaction with Europeans in Muslim Spain.

A third tradition must be mentioned, Arabian and Latin Aristotelianism. It is here only that we can speak of the primacy of reason over faith…. And it so happened that the greatest of the Arabian philosophers, Averroes²[(1126-1198), represented a fairly outspoken rationalism with a frank antitheological bias…Theologians do interpret the (Qur’an), but they do not have the proper training for such an exacting task. The professional philosopher, however are trained to deliver logically necessary arguments and strict demonstrations. They, alone, fulfill the conditions required for a proper interpretation of the Koran. They, alone, are fit to serve as the final arbiter in conflict which may arise between reason and revelation.³

Although it originated from outside Christianity, the third perspective on the relationship between faith and reason nonetheless influenced Christianity.

Aquinas’ Response: The Fourth Solution

Aquinas was well aware of all three trends of approaches to the relationship between faith and reason. However, he was dissatisfied with all of them, and sought to provide a fourth alternative. On the one hand, as a theologian, Aquinas saw great value in faith; on the other hand, Aquinas, the philosopher, saw great value in reason. Since both faith and reason are gifts of God to man, he concluded that they must be both for the benefit of man.

Thomas makes two decisive moves. In the first place, he introduces a fairly consistently maintained separation of philosophy from theology, of knowledge by faith and grace from knowledge by natural cognition.

¹R. A. Markus, Augustine, Critical History of Western Philosophy, 81.
Second sensation – sensory experience – is made the basis of all cognition and knowledge. ¹

By trying to maintain a separation of knowledge obtained by faith and that obtained by reason, one could be misled to conclude that the two means of knowledge lead to different areas of study or to different understandings. The brilliance of Aquinas is in his conclusion that all knowledge is of God, either directly about Him or of His creation, which would indirectly lead to Him. Thus, knowledge of anything at all is knowledge about God.

The distinction between philosophy and theology produces no dualism in the system of Aquinas. The underlying unity of the system is provided, one might say, by the object of knowledge, which is the same in philosophy and theology. All knowledge is knowledge of God. “All conscious things know God in everything they know”. ²

The difference between the two sources of knowledge, revelation and reason, is not the object of knowledge, but the means by which they lead to an understanding of the object of knowledge. Since Aquinas’ theory of human knowledge is basically Aristotelian, natural cognition is based on sensory experience. Given this, its experience is the experience of the creation, not of the Creator. But, because all creation is an effect of the Creator, it reflects and, indirectly, is knowledge of the Creator, God.

By our natural cognitive powers we can make inferences about God by way of knowledge of things He has created. To describe any piece of created nature is also to describe God in so far as a description of the effects of any x is also a description of x. And there is nothing which is not created by God. The difference, then, on which the distinction is based is primarily a difference in method and in the direction of the cognitive process. ³

Thus, Aquinas concludes that there can be no contradiction between the knowledge of revelation and the knowledge of reason. Although Aquinas concludes that there can be no contradiction, it would be a mistake to conclude that the scope and area of the knowledge of reason is the same as that of the knowledge of revelation. Although it is unthinkable that reason and revelation conflict, they do differ in their methods of obtaining knowledge and in their scope and area. In many areas, they overlap, but, not everything discussed in philosophy is pertinent to a person’s religious ends.

³Ibid. 101
Because not every truth admits of the same mode of manifestation, and "a well-educated man will expect exactness in every class of subject, according as the nature of the thing admits," as is very well remarked by the Philosopher (Eth. Nicom. I, 1094b), we must first show what mode of proof is possible for the truth that we have now before us. The truths that we confess concerning God fall under two modes. Some things true of God are beyond all the competence of human reason, as that God is Three and One. Other things there are to which even human reason can attain, as the existence and unity of God, which philosophers have proved to a demonstration under the guidance of the light of natural reason. That there are points of absolute intelligibility in God altogether beyond the compass of human reason most manifestly appears.1

For Aquinas, some truths of revelation that cannot be discovered by natural reason. The Christian mysteries, for example the trinity, must be accepted based on revelation. But even though Aquinas conceded that here are things which reason cannot know by itself, even here, reason has to agree that these thing fall within the realm of possibility. Thus, reason must agree that these theses are possible if it is to accept them as truth obtained through revelation. Therefore, reason and revelation do not conflict even here since truth whatever its source, is true and is recognized as such by both reason and revelation.

Conclusion

The current dominant secular societies deferential attitude towards the religious ethos confronts religion as the dominance of religion in the past had threaten to relegate philosophy to a trivial role. In the past, medieval philosophers’ saw value in embracing the religious ethos and thus enriching both in the process. The question now is, can the religious ethos remake itself to embrace the challenge of reason, thus making itself relevant to society and in the process enrich itself. This can only be achieve, if the religious ethos reaches deep into itself, revives its dynamism and transcends its’ limiting confessional language and appeal instead to all embracing natural law, and therefore achieve the appeal to a common moral language.

In the past, the success of the acceptance of the religious ethos had besieged Hellenistic philosophy and had almost all but suffocated the burning quest and search for truth by means of intellect alone. In the place of reason, divine revelation alone provided answers to all of our questions. By giving all the answers, God terminated man’s bewildering quest for truth and conclusively ended man’s ignorance. The claims of revealed religions took on an destructive offensive against philosophy. Many questioned if there was any use for the inconclusive philosophical speculations which they believed to have been replaced and made redundant by the decisive religious answers.

---

However, the natural curiosity of some men and their burning desire to know for themselves, could never be suppressed or quenched by simple blind faith. Instead, there arose from the ranks of men of faith who sought to transform philosophy to accommodate their religious ethos attempts to find a place for philosophy. They viewed this endeavor as benefit to both reason and revelation. They argued that the attempt to incorporate reason into revelation is not an embellishment, but rather, a fulfillment of the purpose of both. They argued that the purpose of both reason and revelation is not just right understanding, arriving at truths, but also gaining benefits from these truths by acting upon them. The emotive language of revelation, the argued, works well to compel men to action. However, emotional persuasion unsupported by intellectual acquiescent is fleeting and momentary. Once the emotional commitment is intellectually challenged by new ideas, it fades into the abyss or, worse, resorts to anti-intellectual fundamentalism which may lead to fanatical conclusions, forcing them into an escapism from the challenges of the world into a world and logic of their own.
Crossing over the Hindu-Kush, the military campaigns of Alexander brought two of the world's great ancient civilizations into direct contact, establishing a chain of relation which would lead not only to temporary contact but to the establishment of Greek kingdoms in Indian territory for a period of several centuries. This fact provides for the tantalizing possibility of talking about the intercultural philosophical dialogue of ancient Greece with ancient India, a topic of great possible interest both historically, to understand how two such vastly differing philosophical worlds would have interacted, as well as currently, to provide precedents for contemporary dialogues and comparisons between these traditions.

To begin to explore this field, we must assure ourselves that there was truly intercultural philosophical exchange between India and Greece at this time. The traditional approach has examined the issue through the question of influence, hence of identity of ideas, one way or the other. This method renders the question abstract and a-historical, making it very difficult to judge. It is not in the abstract sphere that the question of ancient Greek knowledge of Indian philosophy can be decided but in the practical and historical sphere where actual philosophers with actual philosophical agendas met other philosophers and argued and negotiated with one another's ideas in the search for truth and meaning. In such a dialogue there is a context of understanding

---

1The tradition arguing influence can be seen reflected in Jones (1824), Schlegel (1837) and Chapekar (1977). The tradition arguing non-influence can be seen to greater and lesser degrees in Bellisaropoulos (2001), Banerjee (1981), Sedlar (1980), and Long and Sedley (1987). This second opinion would seem to form the unspoken consensus among scholars. Kartunnen’s (1989, 1997) careful work ends, it seems, in an agnosticism on the issue.
and two interested interlocutors. The negotiation of these two interlocutors according to their interests and within the conditions of the context forms the conditions of the possibility of knowledge. Alexander’s quest for divinization, as an enactment and problematization of the classical ideal of the self formed the context of this encounter. The philosopher companions of Alexander and their debates over this self formed the first interlocutor; Vedic thinkers and conceptions of the self formed the other interlocutor.

This paper will look at the limits but also the possibilities these variables imposed on the extent of possible knowledge of Indian philosophy by Greek philosophers, and affirms that the vision we are given of the Indian philosophers in the histories of Alexander indicate that the philosopher companions of Alexander did engage in philosophical exchange with their Indian counterparts, likely Vedic priests and ascetics, and came to a limited but real knowledge of the Vedic conception of self.

**Enacting the Classic Ideals of Self: The Context of Encounter**

Alexander’s actions in challenging the classical conception of the ethical self, its relation to the divine and freedom formed a context for the encounter of Greek and Indian philosophy. Alexander's campaign took place under the slogan 'Freedom of the Greeks'. While this was firstly a political catch-phrase, its implications were much wider. In the ethical realm, the ‘Freedom of the Greeks’ had to do with the ideal of the Greek citizen as a self-deciding individual within an ethical community of equal selves. At the individual level, ethical freedom was the realization of ethical responsibility for one's actions and the intellectual responsibility to organize and justify them oneself. At the interpersonal level, ethical freedom was the disposition of the self free from the rule of others but according to a code of conduct which would respect and guarantee this freedom not only for oneself but for the other. This was a growth of humanistic values and powers which enabled citizens to engage in actions and practices unthinkable in a traditional, hierarchical and religious society.

In so enabling the individual and empowering him to create his own ethical structure and values, the ‘Freedom of the Greeks’ likewise traumatized the relation between the self and the divine. The growth of the freedom of the ethical self leads to a differentiation and an alienation between man and the divine. The freedom of the self implies the possibility of the independence of the self, its absolute self-deciding. No longer is one’s fate and duty unproblematically determined by the gods. The theories and actions of the sophists, politicians and city-states in the practical realm underlined concretely the decoupling of the morality of the individual and city-state from traditional morality.

It was Socrates who brought this ethical crisis to a head through his interminable questioning of the values of his fellow citizens. On the one hand, he insisted on using his logos, the very instrument of the freedom of the self which enabled the citizen to take this difference from the traditional and the

---

1 Diodorus Siculus.XVII.24.1-2
divine, *ad nauseam*. On the other hand, he took up his constant critique at the behest of a divine voice which left him unsatisfied with the finitude of ethical beliefs and creations of his fellow citizens. Their ethical values, traditional or sophistic, he could always prove to be without ground. Socrates challenged his contemporaries to prove that the freedom of the self is not finally an infinitely groundless turning of reason but can adhere to reason’s rules while respecting the call of the divine to come to a completed and grounded state of freedom of the self. The great ethical philosophies of Plato and Aristotle are both, in their individual ways, attempts to respond to Socrates' ethical challenge.

The philosopher king as both the ruler of an ideal city-state and as ruler of his own heart, represents the revolution of values according to a divine model, the recreation of the world of the given in its incomplete and unsatisfactory nature, according to the model of a perfect world in which reason rules with passion as its co-ruler over the dumb insistence of reality. ¹ The divine is identified with reason. Man is yoked to the divine as the bearer of reason. Thus, the reunification of the free self with the divine is possible and alienation can be overcome.

Aristotle's middle and temperate man are not so much a revolution of values as their reformation. Values are found within the community of free individuals who have a sense of the good. ² The good is a given to be enacted, not a truth to be had. ³ It is thanks to freedom that the individual can recognize this moral truth at all. Human being participates in divinity inasmuch as he can recognize the good in a given situation and enact it. ⁴ The perfect individual, a temperate individual in the highest sense, approximates to the divine through finding the perfect action according to each situation, thus becoming as free as humanly possible.

Classical ethical philosophy reconciles the individual to the divine, by enacting the divine in the world and making this world the world of its freedom thus solving an inherent difficulty within the formula of the 'Freedom of the Greeks'.

Alexander takes up these models of the free self as divine in his quest for deification. This quest is not a simple megalomania but a philosophical quest, grounded in his education. Classical philosophy points the way, if but ideally, towards the possibility of the reunification of man with the divine. The ideal which the philosophers espoused is the ideal which Alexander attempted to accomplish.

Alexander attempted to present himself as a divine and transcendent being. He began in the Greek tradition by tracing his lineage back to Achilles and Heracles. ⁵ He followed in the footsteps of these great heroes, even throwing his spear into the bank of the Asian shores as he first arrived claiming it as 'spear won territory.' ⁶ His campaign is famous for its oracular reports of his divinity.

¹ Plato, *Republic*, 441d-442e
² Aristotle, *Nicomachean Ethics*, 1095b1-10
³ Aristotle, *Nicomachean Ethics*, 1105b13-16
⁴ Aristotle, *Nicomachean Ethics*, 1106B6-9
⁵ Diodorus Siculus, XVII.1
⁶ Diodorus Siculus, XVII.17
On the coast of Pamphylia the seas were said to bow to him, when he successfully negotiated the almost impossible coast line. The most prominent of his oracular portents was his welcome at the oracle of Siwa where he was greeted as the son of the god Zeus-Ammon, and said to be king of the entire world. Alexander's efforts to meet and beat the gods at their own game are a leitmotif of his reign. It is likely that he entered India simply because he wished not to be outdone by Heracles and Dionysus in having conquered India. That Alexander took this divinity seriously is attested to by his efforts to have it recognized not just by his Persian and other Asia troops but by the Greeks and Macedonian troops as well. Alexander offered himself as a model of the philosopher-king and of the temperate man incarnate. He realized the highest dreams of an Aristotle, to unite the Greeks and rule the world, a divine entity above the judgment of his peers, an ethical class of his own.

Alexander's effort to realize the classical ideals of selfhood presented the grounds for a new debate on the nature of self. This debate began even during his own rule and would last well after

The Philosopher Companions on the Self: The First Interlocutor

Alexander’s actions and claims were not uncontested, even within his closest circle. His company included a spectrum of philosophers. These thinkers had been brought into the campaign under the aspect of advisors, researchers, entertainers and even skilled craftsmen. In the company of the king, their arguments on the nature and value of the changes which Alexander wrought on the concept of self are recorded both directly and indirectly. The figures of greatest philosophical stature are four: Callisthenes the Aristotelian, Anaxarchus the Atomist, Onesicritus the Cynic and Pyrrho the Sceptic. The debates and opinions of these men, or at least of their ideas shaped the receptive climate in which Indian thought was taken up during the campaigns of Alexander. The basic poles of the debate over the divinity of Alexander are set out by Callisthenes and Anaxarchus.

Callisthenes contests Alexander's exercise of his freedom of self, his assumption of the mantle of divinity. He must well have realized and approved of Alexander's flirtation with this role in his early years in the campaign but he finds more and more reason to object as the campaign continues. Callisthenes’ preoccupation is to defend the fine line that Aristotle established in allowing for the possible practical existence of the ideal of the

---

1Plutarch, *Alexander* XVII.3; Arrian, *Anabasis*.I.26.1
2Plutarch, *Alexander*. XXVI.5-6; Arrian, *Anabasis*.III.3-4; Diodorus Siculus, XVII.49-51
3By way of example see Diodorus Siculus, XVII.85 and Strabo, *Geography*.XV.1.8 for the storming of Aornus and the challenge to Heracles.
4Plutarch, *Alexander*. XLV
5Aristotle, *Politics*, 1327b29-35
6Aristotle, *Politics*, 1288a15-20
7Callisthenes had been executed shortly before the company embarked for India.
8The fullest account is given in Arrian, *Anabasis*.IV.10.
9On Callisthenes’ role as official propagandist and mythologist of Alexander see Pearson (1960).
great man\textsuperscript{1} while arguing that finally a more moderate form of government must be found. Alexander has surpassed the moderating limits of good government not only of his kingdom but of his self in declaring his divinity in this life. Callisthenes is not willing to cede to Alexander the status of the incarnate divine ethical agent. The great man's free soul must still accord with the teachings of reason itself. Freedom of the soul is the realization of and respect of the ultimate principles as possible within the concrete circumstance. Freedom is not absolute. The Greeks can only be ‘god-like’ because they would never accept a god in their midst. The freedom of the soul is realized in a community, in a given time and place of other free individuals. Without these preconditions freedom of soul can make no sense.

Callisthenes' rejection of Alexander's claim to divinity is brought on by Anaxarchus encouragement of it.\textsuperscript{2} Anaxarchus takes on the unofficial role of Alexander's propagandist in the latter half of his campaigns and follows a radically different intellectual agenda than the Aristotelian. His speech is sophistic in character and his metaphysical background is Democritean atomism.\textsuperscript{3} Freedom of the self as Anaxarchus paints it is the imposition of this self upon the flux of being and time. The self is all powerful if it is willing to take up this freedom. The penance that ultimate freedom pays is that the burden for its actions lies entirely on it. It has not chance for pity of remorse neither for others nor itself. Its actions are law itself, as they form the very basis of law.\textsuperscript{4} A transcendent ideal of the divine does not exist for Anaxarchus, indeed the transcendent as such is ruled out. As he argues sarcastically, the worship of a man when he is already dead is not of much use to him.\textsuperscript{5} The end of freedom is freedom itself, here and now, it has no other destination. It perpetuates itself because the agent wills it to be so. In submitting to the moral and ethical rules of others, this freedom disabuses itself of its own responsibility, deludes itself into placing responsibility for its actions on to others.

In reading Onesicritus, the man and his actions, we should take into account two factors. First, that with Alexander cynicism must be reinvented. Diogenes' cynicism was created in opposition to the city-state; Alexander had already undermined its principles.\textsuperscript{6} Without the foil of the city-state, Cynicism either needed a new target of critique or an exemplar of its cause. In the context of a post *nomos*-based society, the cynic philosopher would be at least in part on the side of Anaxarchus in writing encomiums to Alexander's power. The philosopher-king would seem the ideal ruler in a world where only *physis* decides, not because of his understanding of the ideas but because of his natural will to rule and, through his lack of qualms in following his passions,

\textsuperscript{1}I.e: the man of perfect temperance.
\textsuperscript{2}Arrian, \textit{Anabasis}.IV.9.7, IV.10.5-7; Plutarch, \textit{Alexander}.LI
\textsuperscript{3}Diogenes Laertius, \textit{Lives}; vol.II, 58.
\textsuperscript{4}Plutarch. \textit{Alexander}. LII.3-4
\textsuperscript{5}Arrian. \textit{Anabasis}.IV.10 7
\textsuperscript{6}It is entirely correct to say that the city-state continued to exist and thrive for hundreds of years after Alexander but this does not remove the central point that the city-state as the central agent of political organization was demoted and became a secondary power in the shaping of international politics and therefore of internal politics.
his capability to effectively accomplish this end. Alexander then would have served and did serve as an ideal of cynic kingship. But the exemplar of the cynic ideal could also then be its target of critique. Alexander mixes with his concept of self, a notion of personal divinity that rubs against the deflationist metaphysics of self of cynicism.

Pyrrho's voice in the debate over Alexander's freedom of self was heard passively, reactively but it began to spell out another position on freedom. In all the incarnations and positions it took on around him freedom was distorted and skewed. Freedom embodied in any one doctrine could not ultimately be defended. It would turn into its reverse. Pyrrho in the midst of this argument began to map out the ultimate sceptical position, the position of non-determination, as the position of freedom. Freedom is neither negative nor positive, not in a 'power over' nor a 'power to', but in an abstraction from power as such, in a removal from the entire question.

The Vedic Ideal of Self: the Second Interlocutor

The Indian philosophic tradition is profoundly concerned with the question of self and its relation to the divine from the reflections of the Upanisadic (c.700 B.C.) forward.

Knowing that Alexander's company encountered Indian philosophers but not which, we must determine with which philosophical tradition(s) they came into contact. Although Buddhism, Jainism and Materialism had been in existence as heterodox sects for some three hundred years, their geographical spread at this moment in time cannot accurately be identified. Vedic thought, however, is supposed to have moved in India from the North to the South, and thus we can safely surmise that its teachings were known and practised even in this remote Northern corner of India. For this reason we will assume that the philosophers with whom the Greeks met were trained in the Vedic tradition.

To say that there is a Vedic or Upanisadic conception of the self would be reductive and deceptive, as there exist multiple interpretations of the self and its relation within the Upanisadic themselves and an even greater number of interpretations of these conceptions in commentary upon them. Despite this multiplicity, however, we can identify common traits to the various conceptions of self that form a world picture in strong contrast to the one familiar to the Greeks.

Upanisadic thought conceives of the self (Ātman) in its ultimate form as identical to the creator of the universe as such, a power known as Brahma. This is a fundamentally subject-based idealist vision of reality. This world and all worlds are a result of the projection of the primal self. Worlds and reality are dependent upon the self-projection of the Ātman. Freedom is the realization of one's identity with the primal self and thus the dissolution of arbitrary attachment to temporal and finite worlds, the realization of self in all worlds.

---

1Pearson (1960); Hoïstad (1949), 89-94.
2A good all around view of this interpretation of Pyrrho is found in Conche (1973).
The individual thinking self, lives under the delusion of being other to this one original self. It becomes caught in the worlds of its own projections. It still retains the power of the original Atman, because it is identical with it, but this power is suppressed. Self-realization is the awareness that the world is a projection of the self.

One can be aware of one's nature as world-meaning giver in many ways. As one conceives oneself, so will one's freedom be. One can identify the self with sky, with sun, with air, with space, with the waters or with the earth. Each of these is a true identification of the self but of a limited vision of the self rather, not self in its fullness. The ultimate aim of the Brahman, however, is a total identification,

You seem to know the self of all men in different ways, and you eat food. But the one who worships the self of all men limb by limb, as identified with himself, eats food in all worlds, in all beings, in all selves.

There is a teleology running through the Upanisadic which allows for many different levels of consciousness, but builds towards the ultimate resolution, the absolute freedom that comes from the realization of the single universal self that is not different from the seeking self, Tat tvam asi, ‘You are that.’

The motivation of the path to the absolute is the absolute ontological difference between the freedom of the lesser understanding and the greater. The freedom of a lesser understanding is structurally incomplete. It will lead to rebirths, to suffering and sorrow. The freedom of the greater understanding stands apart.

Where one does not see another, does not hear another, does not know another, that is abundance. Where one sees another, hears another, knows another, that is smallness. Abundance is immortal: smallness is mortal.

The absolute self is self-supporting, without another, an absolute giving that rests only on itself.

This type of comparison is one which is not at all flattering to the worldly accomplishments of men. We find several passages that compare the relative values of the joy accomplished in a perfect human life compared to that of the gods and to those who have achieved spiritual enlightenment. The accomplishments of earthly power such as Alexander’s are compared unfavourably to spiritual accomplishment of the self,

Let there be a young man, a good young man, one who studies, very swift, very steadfast, very strong. Let the whole earth, filled with riches,

---

1Maitri Upanisadic III  
2Chāndogya Upanisadic.V.18  
3See Chāndogya Upanisadic VI.8.7 for use of Tat Tvam Asi.  
4Chāndogya Upanisadic VII.24.1
be his. That is one human joy.\(^1\)

The highest joy of the greatest man is compared with the joy, first of the ancestors, then the gandharvas, the gods, up to Brahma himself. In the \textit{Bhadāra yaka Upani ads}, the joy of the greatest man in a worldly life is equal to one ten millionth of a joy of Brahma, while in the \textit{Taittirīya Upani ads} the figure is one hundred billionth.

The freedom of self of the Vedas is achieved through means very different to those suggested in classical Greek philosophy. There is a rejection of both word and deed as vehicle of achieving this liberation. The difference between world and self is finally rejected and the world is found to be within the self, 

\textit{Subtler than the subtle, greater than the great, The self is hidden in the secret place of a being.}\(^2\) Thus the passionate activity of expanding knowledge and power is seen as futile, useless multiplication.\(^3\) Freedom of self is not far and is not achieved through efforts and activities. Finally, “The person made of all desires, whose distinguishing mark is determination, will and conceit, is bound: the one opposite to this is freed.”\(^4\) Freedom is gained not actively, not through striving but through an openness,

One without will, through the creator's favour,  
Sees the greatness of the self, his sorrow at and end.  
Sitting, it travels far.  
Lying down, it goes everywhere.\(^5\)

Finally, the \textit{Upani ads} have a particular means of passing on the teachings which they embody. It is only late in the tradition that the teachings are written down at all. In the time of Alexander they likely were not. The means of transmitting teachings was oral, between a Guru and his student(s). Because of this there was an ethical responsibility between the teacher and the student, both on the personal and the epistemic level. The inner teachings, the \textit{Upani ads}, are only to be passed on to a certain select few,

In this way a father should teach Brahman to his eldest son or to a trustworthy student, not to anyone else at all, even if someone should give him this whole earth, surrounded by the waters, filled with riches: for this is greater than that.\(^6\)

\(^{1}\) \textit{Taittirīya Upani ads}.II.8.1; See also, \textit{Bhadāra yaka Upani ads}.IV.3.33.  
\(^{2}\) \textit{Ka ha Upani ads}.II.20  
\(^{3}\) \textit{Maitrī Upani ads}.IV.6 “The mind should be kept in check… the rest is multiplication of books.”  
\(^{4}\) \textit{Maitrī Upani ads}.VI.30. See also, \textit{Ka ha Upani ads}.II.9 and \textit{Taittirīya Upani ads}.II.9.1 on the futility of knowledge as means to attainment of enlightenment.  
\(^{5}\) \textit{Ka ha Upani ads}.II.21  
\(^{6}\) \textit{Chāndogya Upani ads}.III.11.5-6; See also, \textit{Bhadāra yaka Upani ads}.VI.3.12 and \textit{Śvetāsvatara Upani ads}.VI.22
The Ideas of the Indian Philosophers of the Greek Histories: This is That

According to the hypothesis that the place of intercultural encounter is not abstract but a practical and interested negotiation, within a given context, the view we should get of Indian philosophy and of its vision of the self should be neither a reflection of Greek thought solely, nor a pure vision of Indian thought in and by itself. Rather it should reflect the contending interests of both the Greek and Indian philosophers within the intellectual context of the debate on the nature of self engendered by the world-conquering, self-divinization of Alexander.

We see in the major encounters of the Alexander histories with the ascetics of the plains (the so-called gymnosophists),\(^1\) with Calanos and Dandamis at Taxila,\(^2\) and with the kings Poros and Taxiles\(^3,\) the Upanisadic notion of self shining through, challenging the self of Alexander and the self of classical Greece. The figures we see reported in the Greek histories which we look at more closely here now, give a fair representation of the basics of Vedic thought on self. The criteria of selection of their ideas represents both a basic knowledge on the part of the philosopher companions of this conception of self.

The philosophers of the plains and the philosophers Calanos and Dandamis both found their freedom in a turning inward to self. The philosophers of the plain make it clear to Alexander that,

> Each man possesses no more earth than the patch we stand on; yet you though a man like other men, except of course that you are restless and presumptuous, are roaming over so wide an area away from what is your own.\(^4\)

There is a double analysis of the self here. First, the self can own no more than the space that contains it, the present body. Second, in attempting to find freedom and exploring the physical world, Alexander distances himself from what is most his own. He has in fact lessened his freedom rather than expanding it, entangling himself in the distraction of externals.

Dandamis’ freedom is similarly based in this inward turning. He argues contra Alexander’s royal decree that his divinity is just as great as that of Alexander’s.\(^5\) Self-possession opens and guarantees a path to the divine. Dandamis’ freedom comes with contentment with and satisfaction with what is given to him: his land, the fruits, the seasons. He is not tempted by Alexander’s gifts, nor by the privations that the king might be able to impose on him. His self is sufficient and inviolable in accord with the dictum,

---

\(^{1}\)Arrian, *Anabasis*.VII.1.2  
\(^{2}\)Arrian, *Anabasis*.VII, 2, 2; Strabo, *Geography*. XV. 1.63 and 68  
\(^{3}\)Plutarch, *Alexander*.LIX-LX.  
\(^{4}\) Arrian, *Anabasis*.VII.1.5-6  
\(^{5}\)Arrian, *Anabasis*.VII.2.3
Fullness comes forth from fullness:
When fullness is taken from fullness,
Fullness remains.\(^1\)

This fullness of the self is verified again in the story of Taxiles in Plutarch, who invites Alexander not to a battle of weapons but of gifts, to show who is a higher man, who can give away most.\(^2\)

Critique of the efforts of Alexander and his men to gain profit and freedom from military adventure and conquest of knowledge are made with every assertion of the Vedic self. The ascetics of the plains say that Alexander’s has set off on a pointless quest, ‘Giving no rest to yourself or others.’\(^3\) Dandamis makes another critique saying, ‘He saw, moreover, that Alexander’s companions were wandering about over all that land and sea to no profit, and that there was no limit to their many wanderings.’\(^4\)

That the lifestyle of the Indian philosophers is reported so persistently speaks to the fact that it was recognized that this was itself crucial to the teaching which these men professed. We even see different stages of Vedic life recorded. The philosophers of the plains speak to the traditions of asceticism in Vedism, not unlikely portraying the life of the forest-dwellers and renouncers or the third and fourth order of Brahmamic life. In the Calanos and Dandamis pair, it seems we see represented the tradition of the Guru and his student (Brahmacārin), the common form of training in Vedic philosophy.

Dandamis, clearly a character representing a higher teacher that the company met is true to the injunction against teaching the inner secrets of Brahma to one who is not ready. He refuses the call of Alexander to join him and teach him directly.\(^5\) Calanos is said to insist to Oniscritus that if he wishes to be taught he ought to strip down like the ascetics and lie on the stones to hear the teachings.\(^6\) When Dandamis actually deigns to teach he does so with a large caveat saying that his knowledge will only be vaguely and approximately understood. Too many barriers stand in the way of understanding not least of which language.\(^7\) Even the fact that Calanos is willing to accept to join Alexander seems in line with Vedic teachings. The younger Calanos could be finishing his time as a Brahmacārin and be coming to the time of the householder, where it is a duty to establish a house and a family.

We can see that the self that the Vedic philosophers present is one which can be of use to the non-classical Greek factions in their arguments over the nature of self. The Cynic can support his rejection of society and his ascetic practice of the control of the body. The nihilist-constructionist can support a theory of a malleable, relative world that is correlate to the will of the individual who is willing to embrace this power. The Pyrrhonian sceptic has its examples of the separation of self from world, knowledge and power, its

---

\(^1\) Prayer recited at the opening of the B hadāra yaka Upani ads.
\(^2\) Plutarch. Alexander. IX.1-2
\(^3\) Arrian, Anabasis. VII.1.
\(^4\) Arrian, Anabasis. VII.2.3
\(^5\) Arrian, Anabasis. VII.2.2; Strabo, Geography. XV.1.63 and 68
\(^6\) Strabo, Geography. XV.1.64
\(^7\) Strabo, Geography. XV.1.64
existence in an indeterminate space. Indian philosophy is taken hold of and used to put the final nail in the coffin of pure classical ethics of self.

Conclusion

Going beyond the question of identity and influence, to look at the concrete circumstances of intercultural exchange between ancient India and Greece we can see that knowledge of Indian philosophy was acquired by Greek philosophers in the time of Alexander. The topic of interest to the philosopher companions of Alexander was the question of self, its relation to the divine and freedom. The knowledge that is taken back from India with regards to Indian philosophy is precisely on this topic. We do not get a full or scholarly picture of Indian philosophy, but one which reflects Greek concern in the question of self and genuinely Vedic answers to this question. Having established that there was a definite knowledge of the Vedic conception of self amongst the companions of Alexander, the next question would be to ask would be to what extent this idea was employed and taken up in later Greek philosophy. Did this knowledge of Indian philosophy have an impact on the formation of Hellenistic philosophies and their overall direction?

Bibliography

Vasubandhu’s use of Vijñapti-mātra: A Reinterpretation in the Context of the Bodhisattva Ideal

Hazel Collinson, The University of Manchester, UK

The fourth century Yogācāra Buddhist thinker Vasubandhu opens his Viś śātikā (Twenty Verses) with the following statement:

“All this is vijñapti-mātra, because of the manifestation of non-existent objects, in the same way as a net of hair may be perceived by someone afflicted with an optical disorder.’1

This is in fact one of the few instances in which Vasubandhu actually uses the phrase vijñapti-mātra. Nonetheless, this verse, and in particular the phrase vijñapti-mātra, has been widely discussed among contemporary scholars of Vasubandhu. In this paper I will attempt to examine this phrase within the wider context of Vasubandhu’s works. Contemporary discussions on Vasubandhu’s use of vijñapti-mātra appear to fall into two quite distinct categories, those that treat it as a kind of philosophical summary of Vasubandhu’s position, and those that examine the phrase primarily in the context of meditation. Notable among the former are Saam Trivedi, Matthew Kapstein, Yoel Hoffmann and Janice Dean Willis, while those such as Stefan Anacker and Paul Williams lean towards the latter position.2

1Viś śātikā 1: vijñaptimātramvedamasadarthāvabhāsanāt | yadvattaimirikasyāsatkeṣo ukā didarśāna ||
2See the following works:
In this paper, however, I will suggest an alternative approach for understanding \textit{vijñapti-mātra} in the works of Vasubandhu. I will endeavour to explore the concept in light of the \textit{Mahāyāna} ideal of the \textit{bodhisattva}. When explored in this context, I hope to show that \textit{vijñapti-mātra} has a dual function. It is both an exposition of Vasubandhu’s philosophical position that there is a conscious process which constructs the world, and also a pedagogical device employed in the context of meditative practice for an ultimately soteriological purpose.

This paper will be divided into three main sections. In the first section I will briefly explore the \textit{Mahāyāna} ideal of the \textit{bodhisattva}. In the second section I will examine Vasubandhu’s use of \textit{vijñapti-mātra} and in doing so I will try to demonstrate that \textit{vijñapti} designates a conscious process which actively constructs our \textit{experience} of an external world. In the third section I will try to show how Vasubandhu’s use of the analogy from dreaming supports this interpretation of \textit{vijñapti-mātra}. I will attempt to demonstrate that \textit{vijñapti-mātra} functions both in a philosophical and meditative context, and that the \textit{Mahāyāna} ideal of the \textit{bodhisattva} provides a link between these two contexts.

Before turning to an exploration of the \textit{bodhisattva} model, I wish to make a brief point concerning the translation of the phrase \textit{vijñapti-mātra}. The second part of the compound, \textit{mātra}, does not really pose a problem in terms of translation. In fourth century Buddhist philosophical discourse, \textit{mātra} came to be used as a technical term meaning ‘only’, and it seems that in the context of the phrase \textit{vijñapti-mātra}, ‘only’ is an appropriate translation. It is also possible that the earlier original meaning of \textit{mātra} has implications for the way in which Vasubandhu employs the term, and this is an idea to which I shall return subsequently. The first part of the compound, \textit{vijñapti}, is rather more problematic. It has been translated in numerous different ways by contemporary scholars. Notable among the English terms used to translate \textit{vijñapti} are ‘ideation’, ‘mind’, ‘perception’ and ‘consciousness.’ For the purpose of this paper I have refrained from translating \textit{vijñapti} at all, in the hope that this might allow for a clearer understanding of the term itself. I will now turn to a brief exploration of the \textit{bodhisattva} ideal.

Saam Trivedi, ‘Idealism and Yogācāra Buddhism’, \textit{Asian Philosophy}, 15.3 (Nov. 2005), 231-246
Matthew Kapstein, ‘Mereological Considerations in Vasubandhu’s “Proof of Idealism” (Vijñaptimātratāsiddhi )’, \textit{Idealistic Studies}, 18 (1988), 32-54
Janice Dean Willis, \textit{On Knowing Reality} (Delhi: Motilal Banarsidass, 1979)
Stefan Anacker, \textit{Seven Works of Vasubandhu} (Delhi: Motilal Banarsidass, 1984)
Paul Williams, \textit{Mahāyāna Buddhism} (London: Routledge, 1989)
The Bodhisattva Ideal

It is perhaps somewhat surprising that discussions of vijñapti-mātra in the works of Vasubandhu have paid relatively little attention to the ideal of the bodhisattva, given that it is so central to Mahāyāna Buddhist thought. It is widely recognised that according to the Mahāyāna tradition there are three types of buddha, of which the bodhisattva is the third. The first type of buddha is the pratyeka-buddha (‘solitary’ or ‘private’ buddha). The second type is the śrāvaka (‘hearer’) buddha. Although both of these types of buddha had attained nirvāṇa, according to the Mahāyāna tradition the path by which they did so was essentially inward-looking, and thus they did not represent the highest ideal. The highest ideal was that of the bodhisattva who, having been released from suffering through the attainment of nirvāṇa, then turned his attention towards alleviating the suffering of others.

This idea is central to the La kāvātāra sūtra, an important foundational text of the Mahāyāna tradition. The following passage is taken from D.T. Suzuki’s translation of the text:

Going over to all the Buddha-lands and assemblages, the Bodhisattvas will listen to the Buddhas, discourse on the nature of all things, which are like a vision, a dream, an illusion, a reflection, and the lunar vision in water, and which have nothing to do with birth and death, eternality and extinction; the Bodhisattvas, thus facing the Tathagatas, will listen to their discourses on the truth that does not belong to the Śrāvaka-and Pratyekabuddha-vehicle.

The significance of the analogies drawn here will be discussed subsequently when I explore Vasubandhu’s use of the analogy from dreaming. At present, however, I merely wish to indicate the value placed on the bodhisattva ideal in the Mahāyāna tradition, particularly in comparison to the śrāvaka and pratyeka buddhas.

That Vasubandhu himself adheres to this ideal is evident in his frequent references to the bodhisattva in the Madhyānta-vibhāga-bhāya, or Commentary on the Separation of the Middle from Extremes. In Chapter Four of this text, for example, he declares that the śrāvaka and pratyeka buddhas

---

1 The Sanskrit term buddha is usually translated into English as ‘enlightened one’ or ‘one who knows.’

2 The literal meaning of the Sanskrit term nirvāṇa is ‘extinguishing’ or ‘blowing out.’ See Rupert Gethin, The Foundations of Buddhism (Oxford: Oxford University Press, 1998), p. 75. Although its usage varies among Buddhist traditions, nirvāṇa can broadly be said to denote the attainment of liberation from the suffering that characterises worldly existence. In the case of Vasubandhu the term can be more specifically understood as liberation from the suffering caused by the misconception that the world exists outside of our construction of it.


The term Tathāgatha literally means ‘the thus gone’ or ‘the thus come’. In Buddhist traditions it is commonly used as alternative word for a buddha. In the above passage it seems that Tathāgata and Buddha are being used synonymously.

---

Vasubandhu’s use of Vījñapti-mātra: A Reinterpretation in the Context of the Bodhisattva Ideal
are essentially inward-looking, meditating upon their own bodies for the sole purpose of becoming detached from them. The bodhisattva, on the other hand, having cultivated mindfulness to the extent of the śrāvaka and pratyeka buddhas, turns his attention outwards, in order to facilitate the cultivation of mindfulness in others.1

According to the Mahāyāna tradition, one of the central ways in which the bodhisattva demonstrates this concern for others is by employing the method of upāyakauśalya, ‘skill in means.’ The La kāvatāra s tra identifies upāyakauśalya as the process by which the bodhisattvas ‘give out varieties of teachings that are based on discrimination.’2 Thus broadly the term can be said to denote the use of linguistic constructs for pedagogical purposes. The practice of upāyakauśalya, however, requires that the bodhisattva remains detached from his own teachings, and furthermore from the very notion of non-attachment itself. The bodhisattva does not think, ‘I am not attached’, for that very thought itself would imply attachment to the notion of ‘non-attachment.’3

This point is emphasised by Edward Hamlin in his article, ‘Discourse in the La kāvatāra-s tra’. Although – as is evident from its title – this article is concerned primarily with the La kāvatāra-s tra, I suggest that Hamlin’s understanding of philosophical discourse in the process of upāyakauśalya can be usefully applied to vijñapti-mātra in the works of Vasubandhu. The following observation is particularly pertinent. Hamlin writes:

…discourse can serve as a vital function in dismantling wrong views and cultivating ‘right understanding’. But in order for it to do so it must never be taken as an end in itself. This is why the Buddha figure in the present text [the La kāvatāra s tra] persistently empties his own logical constructs, just when his interlocutors had begun to place faith in them. In short, discourse becomes skilful means when it doubles back on itself to destructure its own structures.4

I suggest that the same can be said of Vasubandhu’s concept of vijñapti-mātra when viewed in the context of the bodhisattva ideal. It is a pedagogical device employed by the bodhisattva through the application of ‘skill in means.’ In this sense, although vijñapti-mātra relies upon discrimination for pedagogical purposes, the bodhisattva who employs the concept in this way remains detached from it, knowing it to be a linguistic construct. Furthermore, the bodhisattva does not think that he is detached from the concept of vijñapti-mātra.

The bodhisattva both transcends and remains within the world. Gadjin M. Nagao explains this as a process of ‘ascending’ and ‘descending’. He writes:

---

1Madhyānta-vibhāga-bhāya IV.12b, transl. Anacker, in Stefan Anacker, Seven Works of Vasubandhu, p.252
2La kāvatāra s tra 104.120-121, in Daisetz Teitaro Suzuki, The Lankavatara Sutra, p. 104
3See La kāvatāra s tra 104.120-121, in Daisetz Teitaro Suzuki, The Lankavatara Sutra, p. 104
Vasubandhu’s use of Vijñapti-mātra: 
A Reinterpretation in the Context of the Bodhisattva Ideal

…Bodhisattvas, refusing the bliss of nirvāṇa, come down to this world because of their ‘compassion’. For a Bodhisattva, the ascent of wisdom terminates at the point of nirvāṇa from whence the descent of compassion begins. The Bodhisattva is, therefore, characterised by two activities: ‘going up’ or ‘ascending’ and the other ‘coming down’ or ‘descending’.

I suggest that in the ‘ascending’ and ‘descending’ of the bodhisattva it is possible to discern vijñapti-mātra at work in both the philosophical and meditative domains. Vasubandhu does not explicitly instruct the reader to understand the phrase in these two contexts. Nonetheless, I suggest that the way in which he treats the term suggests implicitly that it should be so. Indeed, Vasubandhu’s refusal to give direct instructions may well be an intentional methodological feature of his writing, an indication perhaps of his motivation to release his reader from the grip of conceptual language. In order to begin his ‘ascent’, the bodhisattva is required to participate in philosophical argumentation and analysis. This is evident in Vasubandhu’s use of the analogy from dreaming, which I will explore in further detail below. The analogy offers an explanation of vijñapti-mātra in the context of philosophical discourse, and indeed understanding and participating in such discourse is a necessary preparatory stage on the Yogācāra Buddhist path.

However, the ability to engage with and understand vijñapti-mātra in the context of philosophical discourse cannot in itself lead to liberation. As with all philosophical discourse, the analogy from dreaming and the phrase which it purports to explain - vijñapti-mātra - are themselves linguistic constructs and as such cannot lead to liberation. For this reason vijñapti-mātra must also be utilised in the context of meditation, as a device employed for the removal of misconceptions created by all linguistic constructs, even vijñapti-mātra itself. As Vasubandhu writes in Tṛīṣikā 26-27:

So long as vijñāna does not remain in the state of vijñapti-mātra
Attachment to the twofold grasping will not cease.
Even with regards the recognition ‘this alone is vijñapti-mātra’
Something is caused to remain in front, and [one] does not remain in only that [the state of vijñapti-mātra].

So we can see from this verse that every recognition – including that of vijñapti-mātra – relies upon ‘twofold grasping’ (grāha-dvaya). That is, every recognition relies on the assumption of a subject-object dichotomy, a dichotomy that does not actually exist outside of the conscious process. Thus in


2Tṛīṣikā 26-27: yādvadvijñaptimātratre vijñāna nāvati hati | grāhadvayasyānusayastavattava vinivartate || vijñaptimātramadagamityapi ha yupalambhata | sthāpayannagrata ki cit tanmātre nāvati hate ||
a sense viññāpti-mātra marks the point at which the limits of conceptual language have been reached. The next stage involves the gradual removal of misconceptions through a process of meditation, a process that is beyond language itself.

In terms of the bodhisattva’s ‘descent’, the progression of viññāpti-mātra from the realm of philosophical discourse to that of meditative practice is repeated once more. The bodhisattva utilises viññāpti-mātra as a pedagogical device in both of these realms, in order to facilitate the attainment of liberation in others. Thus viññāpti-mātra plays a vital role in the continual ‘ascending’ and ‘descending’ of the bodhisattva.

It is important, however, that we do not disregard the meaning of viññāpti-mātra in its linguistic setting, even if as a linguistic construct it is ultimately to be relinquished by the Buddhist practitioner. The meaning of viññāpti-mātra as a linguistic construct can tell us much about the way in which Vasubandhu uses the phrase, and so it is to a closer examination of the viññāpti-mātra that I will now turn.

The Meaning of ‘Vijñāpti-mātra’

In this section I will suggest that for Vasubandhu viññāpti designates an active process which constructs our experience of an external world. For Vasubandhu the external world does not exist outside of the process implied by the term viññāpti, but viññāpti still remains. That is not to say that viññāpti ‘exists’ in a naïve realist sense of ‘existence.’ It seems that for Vasubandhu such a notion of ‘existence’ is superfluous. Rather, all that we can be certain of is viññāpti-mātra. Everything must be explained through viññāpti-mātra because despite the appearance of external objects, such objects do not in themselves exist. As Vasubandhu asserts in Viśatikā 1, which I translated in the first paragraph of this paper, ‘objects’ (artha) are ‘unreal’ or ‘non-existent’ (a-sat). In this section I will carry out a closer examination of the phrase viññāpti-mātra in the hope that this might reveal further details about Vasubandhu’s view of the conscious process responsible for the apparent manifestation of an external world. I will first explore Vasubandhu’s specific use of the term viññāpti, and will then look at the second part of the compound, mātra.

The first line of Vasubandhu’s auto-commentary which precedes Viśatikā 1 is almost identical to a passage from the Daśa-bhāma-s tra, (“S tra on the ten Stages”), a key Mahāyāna text. In fact, in the very next sentence Vasubandhu cites the exact quotation from the Daśa-bhāma-s tra.1 He writes, citta-mātram bho jinap trā yaduta traidhātukamiti s trā, ‘From the sātra: the three worlds are (that is to say), oh sons of Jina, citta-mātra.’ In Viśatikā 1, however, rather than use the term citta-mātra, Vasubandhu instead uses the term viññāpti-mātra. The close proximity of Viśatikā 1 to the actual s tra passage as quoted in the auto-commentary

---

1Although Vasubandhu does not give the name of the s tra from which he is quoting, Stefan Anacker identifies the quote as belonging to the Daśa-bhāma-s tra in his translation of the Viśatikā. See Stefan Anacker, Seven Works of Vasubandhu, p. 161
indicates that Vasubandhu makes no effort to conceal this subtle change, which in turn suggests that the use of vijñapti rather than citta must be a deliberate move on Vasubandhu’s part.1 If this is the case, then what are the reasons for this substitution and what do they tell us about how we might interpret vijñapti-mātra in the works of Vasubandhu?

In this context it is useful to identify an observation made by Bruce Cameron Hall, who notes that vijñapti is a noun of action which derives from the causative stem of the verbal root jñā, meaning ‘to know’. Hall observes that in Sarvāstivādin Abhidharma thought, vijñapti was a technical term relating to the three types of karma: mental, vocal and bodily.2 Vocal and bodily karma were classified as vijñapti-karma (‘manifest acts’) and avijñapti-karma (‘unmanifest acts’), terms which denoted whether actions were manifest or not manifest to consciousness.3 For Vasubandhu, as Hall notes, ‘vijñapti...signifies more than vijñapti-karma, but retains a sense of activity or function. Vijñapti designates the basic phenomenon of conscious experience, without requiring its separation into object, subject, and act of cognition.’4 So if we agree with Hall – and I suggest that we do – then it would seem that part of the reason for Vasubandhu’s use of vijñapti instead of citta is to retain this active sense of vijñapti.

But why should we assume that Vasubandhu’s use of the term should be viewed in accordance with this earlier sense? After all, the Yogācāra tradition of which Vasubandhu was a part differed in many respects from the Abhidharma tradition. Nonetheless it is important to realise there are ways in which the Yogācāra tradition demonstrates continuity with Abhidharma thought. This is particularly apparent in their respective views of ‘consciousness.’ Rupert Gethin makes the following observation:

Much of Madhyamaka and Yogācāra theory represents a continuation of and development of particular lines of thought within the broad Abhidharma tradition…the broad Abhidharma framework remains as the theoretical basis for understanding the workings of consciousness, for analysing progress along the path, and for breaking down our basic attachment to self.5

Because Vasubandhu does use the term vijñapti instead of citta, we must assume that there is a reason for this. Vasubandhu has the Yogācāra sūtra material at hand, and could have quite easily restated it word for word. But he does not. One possibility here is that the substitution is made simply on the grounds that vijñapti sits more easily within the metre of the śloka (‘stanza’ or

---

1It is also possible that Vasubandhu is simply using vijñapti as a synonym since the use of synonyms was a common feature of Buddhist philosophical writings in the time of Vasubandhu.
3Bruce Cameron Hall, ‘The Meaning of Vījñāpti in Vasubandhu’s concept of Mind’, p. 9
4Bruce Cameron Hall, ‘The Meaning of Vījñāpti in Vasubandhu’s concept of Mind’, p. 14
verse’). However, it is also quite possible that Vasubandhu’s specific use of vijñapti demonstrates continuity with the Abhidharma sense of the term, however slight this continuation might be.

Although Vasubandhu uses the term vijñapti instead of citta, in the passage of auto-commentary which follows Vi śatikā I he asserts that vijñapti is in fact synonymous with citta, manas and vijñāna. In doing so, Vasubandhu clearly has something important to tell us. But what is it and how does it affect our search for a clearer understanding of vijñapti-mātra? I suggest that in asserting that these terms are synonymous, Vasubandhu is trying to show that the process indicated by vijñapti - or indeed by any of its synonyms – is the only means by which we can have access to the apparently ‘external’ world. This world that appears to exist of its own accord is actually contained within this process. Thus by using the phrase vijñapti-mātra, Vasubandhu implies that vijñapti is sufficient in itself to explain one’s entire experience of an ‘external’ world. Moreover, vijñapti alone is responsible for the construction of such a world, and is consequently the only means through which this world can be accessed.

Vasubandhu, however, is faced with a predicament: The Mahāyāna sūtras employ a number of other technical terms, most notably, citta, manas and vijñāna. As part of the Yogācāra scholastic tradition, Vasubandhu needed to demonstrate his allegiance to the Mahāyāna sūtras. Yet as I have just tried to show, it is through the specific use of vijñapti that Vasubandhu is able to maintain the active nature of the conscious process. So what is Vasubandhu’s solution to this predicament? His solution, I suggest, is to point out that citta, manas and vijñāna are synonymous with vijñapti, and it is in this way that he is able to incorporate legitimate technical terms of the Mahāyāna tradition within his own philosophical system. He is thus able to demonstrate allegiance to the sūtra material, whilst specifically maintaining that all that actually occurs is vijñapti.

So we can see how in his specific use of the term vijñapti, Vasubandhu is able to claim that our experience of an external world is in fact only a construction of an active conscious process, a process which itself, however, does occur. This idea, I suggest, is confirmed if we examine the second part of the compound, the term mātra. As was discussed previously, in the philosophical context in which Vasubandhu was writing mātra was frequently used to mean ‘only’. It seems that Vasubandhu is using mātra in this context, to put forward the position that everything is ‘vijñapti-only’. However, as Bruce Cameron Hall has noted, the original meaning of mātra is ‘measure’ or ‘extent.’1 I suggest that Vasubandhu’s use of mātra in the compound vijñapti-mātra retains something of this original meaning. Viewed in this way, vijñapti-mātra becomes far more than an assertion of the sole existence of ‘consciousness’. In using the term mātra it is also possible that Vasubandhu wants to show that consciousness is the ‘measure’ of the world. Not only does this support the idea that there is a conscious process, it also supports the idea that this process is active, as is indicated in Vasubandhu’s particular choice of the term vijñapti. If vijñapti is the measure of the apparently existing external

1Bruce Cameron Hall, ‘The meaning of vijñapti in Vasubandhu’s Concept of Mind’, p. 13
world, then it must be continuously and actively responsible for the construction of that world, and also for the eventual dismantling of that construction. The conventional philosophical usage of mātra as ‘only’ affirms that this active process is all that there is.

### The Analogy from Dreaming

The idea that vijñapti designates an active process is supported by Vasubandhu’s allusion to the example of dreaming. In his Viśatikā, the objection is raised that if an object were dependent upon perception, then its location in time and space would be ‘illogical’. The analogy from dreaming is used to refute this objection. Vasubandhu states, deśādinyama siddhi svapnavat – ‘restriction as to place, etc., is demonstrated, as in a dream.’ The opponent replies, svapna iva svapnavat | katham tāvat – ‘How is it so much like a dream?’ Vasubandhu replies with the following explanation:

In a dream, without an object of the senses, something – a bee, a garden, a woman, a man, etc. – is seen only in certain places, and not everywhere. And even there, in that place, it is not seen all the time. In this way, even without an object of the senses, restriction as to place and time is demonstrated.¹

In using this analogy Vasubandhu is attempting to show that it is possible to experience an object as though it exists at a particular time, and in a particular place, when the object does not actually exist at all. In fact, Vasubandhu argues, this happens every time we dream.

Not only does the analogy from dreaming explain that a conscious process is responsible for the appearance of external objects, it also supports my previous suggestion that it is only through this conscious process that we can experience objects. As Masaaki Hattori has observed in a paper entitled ‘The Dream Simile in Vijñānavāda treatises’:

These objects vanish as soon as a man awakes from sleep, but the fact that there was during the dream a consciousness which produced the objects to be apprehended remains true. The objects seen in a dream are unreal, and the pleasure and the pain experienced in a dream are untrue, but the consciousness in a dream is undeniably real. Likewise, according to the Vijñānavādins, the external objects are unreal, but the consciousness which produces external objects and grasps them really exists.²

---

¹Viśatikā-v tī 3a: svapne vināpyarthe na kvacideva deśe ki cid bhramarārāmastra puru ḍārika d śyate na sarvatra | tatraiva ca deśe kadācidd śyate na sarvakālamiti siddho vināpyarthena deśakālanīyama |
In the context of Vasubandhu it would perhaps be more helpful to say that the conscious process occurs, rather than ‘the consciousness…really exists.’ Nonetheless, Hattori’s point is a particularly pertinent one. Vasubandhu’s use of the analogy from dreaming is not merely a denial of external objects, but an assertion of the conscious process that constructs the appearance of these objects.

Furthermore, I suggest that Vasubandhu’s use of the analogy from dreaming also supports the active sense of viññāpti which I have been arguing for in this paper. To clarify, the conscious process causes the appearance of ‘external’ objects, thereby reinforcing the causative sense of viññāpti. As Edward Hamlin has observed, ‘The elements of a dream do not come and go of their own accord, whatever their appearance to the contrary; they come and go because we make them do so.’ Thus Vasubandhu’s use of the analogy from dreaming reinforces the idea that viññāpti indicates a conscious process responsible for constructing the world as it appears to us.

Moreover, allusion to the example of ‘dreaming’ was a conventional feature of Buddhist philosophical discourse. Thus by using the analogy from dreaming Vasubandhu was able to locate and legitimise his assertion of viññāpti-mātra within a Buddhist intellectual context. ‘Dreaming’ is one of ten analogies identified by Masaaki Hattori in his aforementioned article. Among the nine others are māyā (illusion), mar ci (mirage), udakacandra (the moon in the water) and chāyā (shadow). In using the analogy from dreaming, therefore, Vasubandhu was able to legitimise his place in an ongoing Buddhist scholastic tradition. He takes a legitimate source for philosophical argumentation and embeds it within his own philosophical system, in much the same way as he does with the s tra material.

Conclusion

In this paper, I have attempted to offer an alternative perspective on the phrase viññāpti-mātra in the works of Vasubandhu, through exploring the concept in light of the bodhisattva ideal. I have tried to show that when viewed in the context of the bodhisattva ideal, viññāpti-mātra can be seen to function as a philosophical explanation of the misconception that there is an ‘external’ world, but also as a pedagogical device employed in meditative practice for the purpose of release from this very misconception. An understanding of viññāpti-mātra in its philosophical context is a necessary stage on the bodhisattva path. However, liberation can only be attained though the process of relinquishing all attachments, a process which is only possible through meditation. When the Buddhist practitioner has attained liberation, he turns his attention towards other sentient beings, and the entire process begins once more.

1Edward Hamlin, ‘Discourse in the La kāvatāra-s tra’, p. 277
Vasubandhu’s use of Vijñapti-mātra: 
A Reinterpretation in the Context of the Bodhisattva Ideal

Through viewing vijñapti-mātra in this way I have endeavoured to cut across the distinction between philosophy and meditation which pervades contemporary writings on Vasubandhu. I have attempted to show that when viewed in this context, vijñapti-mātra indicates an active conscious process which constructs our experience of the world as if it were external. As such, it provides a philosophical explanation of why we cling to the false idea of an external world. This can be seen as serving a dual purpose. Firstly, it firmly establishes Vasubandhu’s view on the relationship between consciousness and the world, a necessary and important undertaking in the Sanskritic intellectual climate of which Vasubandhu was a part. It also offers a philosophical explanation as to why sentient beings suffer. They cling to this false construction of an external world, characterised as it is by fictitious dualities, believing it to be truly existent. This tension - between the constructed nature of the world and our conviction of its reality - inevitably leads to suffering.

If we were to understand vijñapti-mātra purely in this explanatory capacity, then the picture that it paints is a bleak one indeed. However, as I hope to have demonstrated in this paper, vijñapti-mātra does more than just explain our attachment to false dualities and the suffering that ensues from this attachment. As a pedagogical device, it also presents a soteriological opportunity, providing a practical means for transcending suffering. This is where vijñapti-mātra ceases to be simply a philosophical concept. In fact, as a philosophical concept vijñapti-mātra is a linguistic construct, and if viewed only in this way would in fact serve to perpetuate suffering. Thus according to Vasubandhu we must eventually relinquish the very concept of vijñapti-mātra itself.

Through meditation upon vijñapti-mātra, the Yogācārin is finally released from the false perception of the world that is the root of suffering. However, although this is the point at which the practitioner has attained arhatship, it is not the final ingredient of the bodhisattva path. The Yogācārin is then driven by compassion to alleviate the suffering of others, and in this way attains the status of the bodhisattva. As has been discussed, one of the central methods employed for this purpose is that of upāyakauśalya - ‘skill in means.’ This entails the use of linguistic concepts for pedagogical purposes. I have attempted to demonstrate that vijñapti-mātra is one such linguistic concept. It explains Vasubandhu’s philosophical position on the conscious process which grants us access to the appearance of a world, which in turn explains why it is that we suffer. However, it also offers a practical way out of this suffering. Just as the Yogācārin returns to the world in the capacity of the bodhisattva, so too vijñapti-mātra is utilised again as a pedagogical device for the purpose of facilitating enlightenment once more.

In this paper I hope to have shown that the way in which Vasubandhu employs vijñapti-mātra actually cuts across the apparent divide between philosophical discourse and meditative practice. In order to clarify this point, it may be useful to refer to another analogy not yet mentioned in this paper. The analogy to which I am referring is found in the following passage from the La kāvatāra sūtra:
As a king or wealthy householder, giving his children various clay-made animals, pleases them and makes them play [with the toys], but later gives them real ones; So, I, making use of various forms and images of these things, instruct my sons; but the limit of reality (bhātaiko) can [only] be realised within oneself.¹

I suggest that not only does this passage emphasise the limitations of conceptual language – by explaining the way analogies are employed by reference to yet another analogy – but it also provides a useful model for understanding Vasubandhu’s use of vijñapti-mātra. It might be useful to understand vijñapti-mātra as a linguistic concept in the same sense as a clay model, provided for the reader to play with on the initial stages of the path to liberation. Eventually, the Buddhist practitioner will relinquish attachment to the clay model. Similarly, as was noted earlier in reference to Triśikā 26-27, even the concept of vijñapti-mātra itself must be relinquished. Once the Buddhist practitioner has reached this point, the clay model is taken up once more, and given to another, in the hope that they too will attain the ideal of the bodhisattva. It is in this sense that Vasubandhu’s understanding of vijñapti-mātra functions not only in the context of philosophical analysis and argumentation, but also as a pedagogical device on the meditative path. By interpreting vijñapti-mātra in this way, I hope to begin to challenge the perceptible divide – between philosophical discourse and meditative practice – that exists in contemporary writings on Vasubandhu.

¹Laṅkāvatārastra 2.147-148, in Daisetz Teitaro Suzuki, The Lankavatara Sutra, pp. 77-78
The Hollow Truth of Ockham’s Razor

Jeffrey Goodman, James Madison University, USA

In this paper I wish to defend Ockham’s Razor from one sort of objection that is often leveled against it. Specifically, I will argue that there is at least one substantial version of the Razor that is true and immune to the objection; when interpreted as a certain normative principle regarding ontological commitment, the Razor counsels correctly.

In discussing some normative principles regarding ontological commitment, Oliver (1996) remarks that the ‘Ockhamite prohibition against an overly generous ontology or, as it is usually put, “do not multiply entities beyond necessity” …seems to boil down to the rather banal exhortation not to believe in the existence of anything unless one has a reason to so believe.’¹ This interpretation is perhaps banal, but whether it has or lacks this feature is not a focus of the present paper. What I first wish to focus on, rather, is an explication and a defense of the following pair of claims that may appear on their face to be incongruous: 1) So understood, Ockham’s Razor is true – at least when it is interpreted not as an imperative but as the following normative claim (which I shall henceforth call \( BE \)): If an agent \( S \) has no reasons whatsoever to believe that \( x \) exists, then \( S \) ought to believe that \( x \) fails to exist. That is, adhering to \( BE \) leads to justified beliefs about the existence of things; it is in this way a useful guide regarding our ontological commitments; 2) \( BE \) is, in an important sense, vacuous.

An objection that often crops up in discussions of Ockham’s Razor runs as follows. On typical construals of the Razor, we have a principle that advocates believing in the particular entities whose existence is entailed by some ontology \( T \) and disbelieving in the entities whose existence is entailed by a rival ontology \( Q \) just in case \( T \) is simpler, more elegant, tidier, less extravagant, etc. than \( Q \). But such features are purely aesthetic; as such, they give us no epistemic grounds for preferring one ontology over its rivals. It is this objection that I here wish to deflect. For now, however, we may just point out that since

**BE** does not make explicit mention of any such features, this sort of objection simply does not apply to it. That is some reason, I suppose, for thinking that **BE** is not a genuine version of the Razor. But I am not very interested in that topic. Rather, I intend to argue below that there is another principle that **BE** bears importantly on that is a genuine version of the Razor, and that this version is true. Before we get to that, however, let us first examine (1) and (2) from above.

First, let us see why **BE** is true. To establish this, we can consider an even wider principle – one that is not restricted to claims of existence

**B**: If an agent S has no reasons whatsoever to believe proposition p, then S ought to (merely) believe not-p (i.e., S ought to (merely) disbelieve p).

There is a natural objection to **B** that immediately arises. The absence of any reason had by S for p does not entail that a belief in not-p is the only justified attitude for S. Rather, in such cases, S simply ought not believe p, i.e., the justified attitude for S is either disbelief in p or suspension of judgment about p.

This objection fails, however. Suppose S has absolutely no reasons for p. That is consistent with S having many reasons against p. In such a case, then, S ought to believe not-p. It would thus be epistemically wrong in such a case for S to suspend judgment about p. But more generally, on the assumption that the antecedent of **B** is satisfied, suspension of judgment could only be a justified attitude for S were she to have absolutely no reasons for not-p. (I think that suspension of judgment is a justified propositional attitude S has toward p if and only if the weight of the reasons S has for p is equals to the weight of the reasons S has for not-p.) But as I shall now argue, that is impossible.

For any proposition p entertained by an agent S, there is always some reason for S reject p. No matter what objective probability is assigned to p, and no matter what subjective probability is assigned to p, S always has some reason for believing not-p, and here is why. Even the most cocksure agent ought to recognize that she is fallible. That is, every agent ought to be aware that she might have been mistaken about some past beliefs, is potentially in error now about some of her beliefs, or will perhaps have some mistaken beliefs in the future. Now, there may indeed be a handful of propositions that, if believed by S, S could not possibly be mistaken about. Perhaps propositions concerning one’s own existence, one’s own mental states, or mathematical truths are like this. But that still does not imply that S has absolutely no evidence against such propositions.1 Every agent (even those non-philosophers unaware of the possibility of the Cartesian Evil Demon, Putnamian Brain-In-A-

---

1We could take a piecemeal approach and examine one by one propositions in which it seems **prima facie** that the agent that believes them has no reason whatsoever to doubt them. Take, for instance, the belief I have that I exist. For me, the fact that a capable philosopher, viz., Peter Unger, has sincerely doubted his own existence gives me some reason to doubt my own. Or consider mathematical truths. For me, the fact that I have added numbers incorrectly in the past gives me some reason to doubt my belief that 2 + 2 = 4. (Of course, these beliefs are justified for me in the end, and the reasons cited for doubting them are epistemic drops in the ocean.) And although it seems incredibly easy to give quite a bit of credence to the claim I’m making here by taking this piecemeal approach for a large number of possible beliefs, it would not be a principled way to support my claim. It would force a piecemeal handling of alleged counterexamples that would rightly seem **ad hoc**.
Vat scenarios, etc.) has the following argument at her disposal for any proposition that she is currently entertaining: ‘I am entertaining p; a number of propositions I have entertained in the past, or will entertain in the future, are ones I might have been or might turn out to be mistaken about; p may be among them.’

So, any rational agent S always has some reason for believing not-p. By parity of reasoning, then, doesn’t that imply that the case imagined – i.e., the case where the antecedent of \( B \) is satisfied – never obtains? That is, doesn’t that imply that there are no cases where an agent S has no reason whatsoever for p? Yes. And it is in this sense that \( B \) is vacuous. For any proposition p entertained by an agent S, there is always some reason for S to believe p. No matter what objective probability is assigned to p, and no matter what subjective probability is assigned to p, S always has some reason for believing p. Even the humblest of agents ought to be aware that she may have been correct about some of her past beliefs, that she may be correct about some of her current beliefs, or that she is capable of some correct future beliefs. Now, there may be a handful of propositions that, if believed by S, S is guaranteed to be mistaken about. Perhaps propositions concerning the existence of one’s own limbs when one is merely a brain in a vat, or mathematical falsehoods, are like that. But that does not imply that S has absolutely no evidence against such propositions. Every agent at least has the following argument at her disposal for any proposition that she is currently entertaining: ‘I am entertaining p; a number of propositions I have entertained in the past or will entertain in the future are ones that I might have been or might be correct about; p may be among them.’

So, the normative principle currently being considered is nowhere and at no time perfectly applicable. The antecedent of \( B \) cannot be satisfied, and it is in this sense vacuous. But \( B \) is not false; it is indeed a good guide for our epistemological practices. If there were, per impossible, some proposition p that an agent S had absolutely no reason for, then S would (only) be justified in believing not-p (and not justified in suspending judgment about p). To gain a firmer grasp on how the epistemic principle \( B \) is useful, let us compare it to the following scientific principle:

\( \text{N: If an object is not acted upon by an outside force, then the object, if in motion, will stay in motion, and if at rest, will stay at rest.} \)

\( \text{N} \) is a version of Newton’s First Law. Of course, we do not live in a Newtonian world, but it is fairly easy to see that were our world Newtonian, and were every object in existence acted upon by some outside force, \( \text{N} \) would be vacuous in the same sense that \( B \) is vacuous. That is, the antecedent of \( \text{N} \) would be nowhere and at no time satisfied by any object. But \( \text{N} \) would nonetheless be true. Moreover, \( \text{N} \) would be useful; it would be a good guide for our scientific practices. There would be objects that were acted on by very few forces, and we could thus use \( \text{N} \) to make reasonable predictions regarding their behavior. Similarly, then, I claim that even though there are no propositions p believed by a rational agent S for which S has absolutely no reason, there are propositions for which S has very few reasons. And \( B \) correctly leads S in the direction of disbelief in such instances.
Returning to BE, we can note that since this principle is simply a restricted version of B, it is true if B is true. And the truth of B has just been established. Hence BE is true.

Now we wish to investigate the following: How does the truth of BE bear on Ockham’s Razor? I stated above that we would examine the charge often leveled against the Razor that features that appear to be purely aesthetic (simplicity, tidiness, elegance, etc.) have nothing to do with the epistemic reasons for accepting an ontology, and hence have nothing to do with whether or not we ought to commit ourselves to the entities the ontology entails. I now wish to deploy the results from above regarding BE to show that this is not a successful criticism.

Consider the following construal of Ockham’s Razor that was suggested earlier:

**OR**: If some ontology T is simpler, tidier, and more elegant than some rival ontology Q, one should believe in the particular entities whose existence is entailed by ontology T and disbelieve in the entities whose existence is entailed by Q.

The results from above tell us that we never in fact have absolutely no reason to believe a proposition we are entertaining, and we never in fact have absolutely no reason to believe its negation. There are just various reasons for and various reasons against any proposition being entertained, and it seems reasonable for an agent S to believe a proposition p iff the totality of the reasons S has for p outweighs the totality of the reasons S has for not-p; likewise, it seems reasonable for S to disbelieve p iff the totality of reasons S has for not-p outweighs the totality of reasons S has for p. (And to reiterate a point from above: it seems reasonable for S to suspend judgment about p when the totality of reasons S has for and against p are weighted equally – it just turns out that the weight of reasons S has on either side of a p never equals zero.) Now, I claim that if the arguments given above for the truth of BE are successful, then the OR detractor is mistaken in thinking that aesthetic features cannot justify a commitment to one ontology over its rivals. The simplicity, tidiness, and elegance of an ontology T (which quantifies over entities (of kind) x) are epistemic reasons for S to prefer T (and its concomitant ontology) over rival ontology Q because they offer some support to S for T of the following variety: ‘T is relatively more simple, tidy and elegant than Q; many past theories that are relatively simpler, tidier and more elegant than their rivals have turned out to be true, while the relatively more profligate, messier, less elegant theories have turned out to be false; so, T may be true.’

What makes OR a substantive and interesting (true) epistemic principle is that these features are often times enough (i.e., jointly sufficient) to outweigh the totality of reasons against some proposed ontology T. That is, there are times when an agent has very few reasons to doubt some ontology T and very few reasons in favor of T other than reasons that appeal to aesthetic virtues. OR rightly counsels us to accept T in such instances. So, the features appealed to may be aesthetic, but these need never be counted as purely aesthetic in cases where they are used as evidence for some ontology in the way discussed
above. Such features are epistemic as well in any such case where they are
taken to be part of all the evidence for (or against) some ontology.
All that has been said here ignores what it is to be a reason, but it obviously
takes the threshold to be low. All that has been said here ignores the distinction
between good and bad reasons, as well as the distinction between powerful and
weak reasons, etc. (i.e., we have ignored the particular weightiness of
individual reasons). All that has been said here assumes that agents are capable
in some sense of weighing the totality of reasons for any proposition under
consideration and the totality of reasons against any proposition under
consideration and calculating which is greater. And all that has been said here
implies very little about the nature of the support relation (but not nothing –
e.g., it clearly implies that the relation is not deductive entailment or even
strong induction). These issues being ignored and the assumptions being made
are, I realize, all controversial and require discussion. I do think that all of the
assumptions made are plausible, though, and that all of the controversial
matters can be settled in principle. However, they are matters that must be left
for future papers and other philosophers.

\footnote{But not completely. They are somewhat like little arguments (inferences) occurring in the
head.}

\footnote{Not to mention the ‘goes against’ relation.}
The primary focus on the epistemology of the *Apology* has for a long period of time been on the concept of ‘knowledge.’ Scholars have questioned whether Socrates disavows knowledge, and whether he does so disingenuously. Perhaps Socrates genuinely disavows knowledge but not true beliefs? Perhaps Socrates only disavows one kind of knowledge – infallible knowledge, or expert knowledge, for example – but not another kind of knowledge – fallible, elenchic knowledge, or the knowledge of a layperson? The analysis of the *Apology* in terms of ‘knowledge’ has taken the focus away from the concept of ‘wisdom.’ This is unfortunate because the distinction which Socrates himself draws in the *Apology* is not between two kinds of knowledge but between two kinds of wisdom: human wisdom and a wisdom that is somehow ‘more than human.’ In this paper I will provide an analysis of the concept of wisdom and the distinction which Socrates draws in terms of it. Firstly, I will argue that the wisdom which Socrates disavows is a *technē*. Secondly, I will show that Socrates’ inquiry into the claim of the oracle can also be read as a kind of epistemological inquiry into the concept of wisdom. I will suggest an interpretation on which Socrates on the basis of his investigation of the politician, the poets, and the craftsmen in the *Apology* concludes that the wisdom which it is possible to possess as a human being is principally to be thought of as self-knowledge and not as *technē*. I will not

---

1 I do not want to beg any questions of when or to what extend - if at all and to any extend – the ‘Socrates’ we find in Plato is the historical Socrates. Nor do I want to beg any questions of whether ‘Socrates’ is nothing more than Plato’s mouthpiece or not. By ‘Socrates’ I simply mean the character ‘Socrates,’ which we find both in the *Apology* and in Plato’s other dialogues.


3 ‘Self-knowledge’ is throughout to be taken as ‘knowing what you do not know.’ I will not discuss how self-knowledge relates to ‘knowing yourself’ or ‘knowledge knowing itself.’ I will not discuss how self-knowledge relates to sōfrosunē either. See Annas (1985), Carone (1998),
translate *technê*. Despite disagreements there is a broad scholarly consensus about the following, which I will assume: *technê* is often used interchangeably with *epistêmê*, and it denotes a consistent body of knowledge that enables its possessor to teach it and to give an explanation of the subject covered by it.¹ In so far as a *technê* denotes a whole body of knowledge it is often translated as ‘expert knowledge’ or ‘expertise,’ and in so far as *epistêmê* does the same it has to be taken in the same thick sense.² The thickness of *epistêmê* and *technê* makes it natural to regard both as ‘wisdom,’ *sofia*, and Socrates often uses different *technai* as prime examples of both knowledge and wisdom.³ It would therefore be remarkable if Socrates suggests that wisdom is not best to be thought of as *technê* in the *Apology*.

**The Wisdom of the Sophists**

At 20d6-e2 Socrates distinguishes between a kind of wisdom he avows - human wisdom - and a wisdom he disavows - a wisdom that is greater than human wisdom:

> I have, Athenians, acquired this reputation on account of nothing other than a sort of wisdom (*sofia*). What kind of wisdom is that? The very one which is perhaps human wisdom (*anthrôpinê sofia*); in truth I probably am indeed wise in that. But perhaps the other men I mentioned just now are wise in a sort of wisdom greater than human – or else I don’t know what to say. I do not have knowledge of it. (*Apol.* 20d6-e2)

What is the wisdom, which Socrates respectively avows and disavows? The disavowal and the reference to ‘the other men I mentioned just now’ is to be seen within the context of a conversation between Socrates and Callias, which Socrates has recounted to the jury:

> ‘Callias,’ I said ‘if your two sons had happened to be a pair of colts or calves we should have been able to put somebody in charge of them and hired a man likely to make them fine and good in the appropriate virtue, and this man would have been either an equestrian or an agriculturist. But as things are, since they are a pair of human beings, whom do you have in mind to put in charge of them? Who is there who is knowledgeable in the appropriate virtue, both that of a human being and that of a citizen?… Is there such a man,’ I asked, ‘or not?’ ‘Certainly,’ he answered. ‘Who’, I said, ‘and where from, and at what price does he teach?’ ‘Evenus, Socrates,’ he said, ‘from Paros, for five minae.’ And I

---

deemed Evenus to be blessed if he truly possessed this technê and teaches it so reasonably. I at least for my part would actually be priding myself and putting on airs if I knew these things, but I do not know them. (Apol 20a7-c3)

Socrates wonders whether there is a technê of human virtue in the same way in which there is a technê of colts and calves. Callias emphatically replies that there is, but Socrates does not seem to be fully convinced, since his praise of the sophist Evenus is conditional. However, he is fully convinced that he himself does not possess a technê of human virtue, and he explicitly disavows knowledge about moral matters. It is clearly this disavowal which is repeated, when Socrates disavows a ‘wisdom that is greater than human,’ and the contrast between ‘human wisdom’ and ‘some wisdom that is greater than human’ is therefore a contrast between human wisdom and the sophists’ alleged technê of human virtue. This contrast, however, is odd. Within the context of 19d9-20c4, the only obvious candidate of human wisdom, anthrôpinê sofia, a phrase which is not self-explanatory and likely to puzzle the reader as he encounters it for the first time, is the sophists’ alleged technê of human virtue - anthrôpinê aretê. The sophists advocate themselves as sophists, sofistai, literally ‘wisdomists,’ because they take themselves to have a special claim to wisdom. They do not teach somebody to become a good carpenter or a good tanner etc., but to become a good human being. And because the virtues, which the sophists teach, are human virtues, and not the different virtues of the different handicrafts, the alleged wisdom of these ‘wisdomists’ is naturally seen to be human wisdom. It is therefore most peculiar that Socrates calls his wisdom, which he contrasts to the wisdom of the sophists, for human wisdom. What would human wisdom, anthrôpinê sofia, be a technê of if not of human virtue, anthrôpinê aretê? Because there is no obvious answer to this question, the reader is left wondering whether human wisdom is a technê at all. Perhaps human wisdom is best thought of as a non-technical kind of wisdom?

Socrates’s distinction contains another puzzle. Socrates calls the wisdom he attributes to the sophists for ‘greater than human,’ but it is immediately puzzling how sophists who, after all, are human, can possess a wisdom that is greater than human. Socrates’ statement could, of course, be read as a straightforward compliment. He is awe-struck by the wisdom of the sophists, and he genuinely takes them to be wise. On the other hand, the claim might also be seen to indirectly suggest that the sophists lay claim to a technê and a wisdom that it is impossible for a human being to possess. If that is the case, Socrates’s attribution of a wisdom which is ‘more than human’ is to be read as cancelling itself out, and thus as a denial that the sophists in fact possess any wisdom at all. The fact that Socrates made his praise of Evenus conditional adds to the suspicion that he does not deem the sophists to be wise.

Human Wisdom

Although Socrates disavows the alleged moral wisdom of the sophists he is happy to lay claim to ‘human wisdom.’ In order to give an account of human
wisdom, Socrates proceeds to tell the tale of Chaerephon’s visit to the oracle. The oracle proclaims that no one is wiser than Socrates. Socrates is puzzled:

What on earth is the god saying, and what on earth lies behind his riddle? In nothing great or small do I know myself to be wise. What on earth, then, is he saying when he claims that I am the wisest? (Apol. 20b3-5)

Socrates has a concept of wisdom, and it is because he does not take himself to fulfil the conditions of it that he cannot understand why the Oracle has pronounced him to be the wisest. Indeed, it appears that Socrates cannot even understand why he should be wise at all – let alone be the wisest – since he denies that he is wise in anything great or small. On what conception of wisdom does Socrates deny that he is wise? And what is the difference between being wise in something great, on the one hand, and in something small, on the other?

Given the evidence of how natural it is to take technê to be wisdom, an obvious answer to the first question is: on a conception of wisdom as technê. As a positive body of expert knowledge technê has a natural claim to wisdom. The answer to the second question is probably also relatively straightforward. There are many kinds of technê. The tanner, the carpenter, the doctor, the pilot, the politician and the sophists all have a claim to wisdom in so far as they really possess a technê. But the degree of importance and prestige of the different technai varies.¹ Tanners, carpenters, doctors etc. cannot unqualifiedly call themselves for sofistai, ‘wisdomists,’ or even for wise, because their wisdom is considered relatively unimportant and non-prestigious, whereas the sophists, the orators, and the politicians can. The greatness and smallness which Socrates refers to is therefore presumably relative to the importance of the different technai. By denying that he is wise in anything great or small Socrates is most naturally understood to be denying both that he possesses any of the menial technai and that he possesses any of the important and prestigious technai such as the technê of virtue.²

Socrates sets out to examine the meaning of the Oracle by trying to find somebody who is wiser than he. First he goes to a politician.³ We do not get an account of Socrates’ investigation of the politician but we can imagine that

---

1See for example Grg. 448c, 452a-453a, and 512c-d.
2This interpretation seem to contradict the suggestion that Socrates, son of a stonecutter, would have learned the technê of stonecutting himself, since the technê of a craftsman, all other things being equal, is passed down from father to son. However, it might be the case that Socrates no longer possessed a technê for lack of practise even if he had learned a technê such as stonecutting in his youth. I.e. even if Socrates had once learned a technê he could still truthfully claim that he does not possess any technê when he was inquiring into the meaning of the Oracle’s saying.
3It is not an accident that Socrates goes to a politician and not, for example, to one of the handicraftsmen first. The politician, as I argued above, has an unqualified claim to wisdom which the handicraftsmen do not have. Socrates coyly proclaims that there is no need to mention the name of the politician (Apol. 21c3). A natural suggestion, since the politician is said to come to dislike Socrates, is that the politician is no other than Anytus, who is one of the three accusers. Anytus also threats Socrates in Meno 94e.
Socrates asks the politician pertinent questions such as ‘what is justice?’,\(^1\) which he is unable to answer to Socrates’ satisfaction - presumably because he gets entangled in inconsistencies in the way in which Meletus is going to get entangled in inconsistencies during Socrates’ investigation of him (\textit{Apol} 27a2-7).\(^2\) Socrates concludes that the politician does not know what he thinks he knows, and that he, Socrates, is wiser than the politician:

> And then I tried to show him that, though he thought he was wise, he wasn’t. So, as a result of this, he and many of the people who were there took a dislike to me; but I thought to myself as I went off that I was wiser than this person at least: probably indeed neither of us knew anything fine and good, but he thought he knew something he didn’t, whereas I, just as I didn’t know, didn’t think I knew. I seemed likely, therefore to be wiser than him by virtue of a small thing, this very point, that what I didn’t know I didn’t think I knew either. (\textit{Apol}. 21c7-d8)

Socrates starts off by being surprised that he is wise. But at the end of his examination of the politician he is no longer surprised that he is wise for he can now give an account of a way in which he is wise – he is wise by virtue of self-knowledge, i.e. by virtue of knowing what he does not know. The question is in which of these two ways Socrates solves the puzzle of his wisdom:

1. Does Socrates maintain his initial concept of wisdom but change his judgements of whether he satisfies the necessary and sufficient conditions?
2. Does Socrates revise his concept of wisdom and therefore also his judgements of whether or not he is wise?

I suggest the latter. Socrates already possessed self-knowledge from the outset of the investigation. Indeed, it was because he already knew that he did not possess any \textit{technē} at all that Socrates was puzzled when the oracle said that no one was wiser than him. If Socrates at this initial stage had thought that wisdom was essentially self-knowledge he would presumably not have been puzzled,\(^3\) so he must have start off with a different conception of wisdom. The suggestion that Socrates changes his concept of wisdom from something to something else is corroborated by the passage quoted above. Socrates tries to show the politician that he is not wise before he formulates his new account of wisdom as self-knowledge. Socrates must therefore have a concept of wisdom, which the politician fails to live up to, that is different from that of self-knowledge.

\(^1\)The politician’s’s claim to wisdom is related to what he is supposed to know about human virtue in general, and justice and injustice in particular (see \textit{Grg}. 459-460a, and \textit{Alc}. 1 106c-109c)

\(^2\)See Burnyeat (1997) for a discussion of whether Socrates thereby successfully rebuts Meletus’ accusation. Burnyeat argues that Socrates succeeds in showing that Meletus is muddled up, but that he does not refute that for which he in fact stands accused – namely for ‘not accepting the gods that the city accepts but in other demonic beings’ (\textit{Apol}. 24b8c1).

\(^3\)Unless he presumed that somebody somehow had more self-knowledge than he did, which we have no reason to suspect.
knowledge (which, as it so happens, the politician also falls short of). ‘Wisdom
is technê,’ is the most likely candidate.

Socrates takes himself to be superior to the poets in the same way in which
he is superior to the politician (Apol. 22c6-8), and he finally turns to an
investigation of the craftsmen.

Finally, then, I came to the craftsmen; this, because I was conscious of
knowing, myself, almost nothing, but I knew that I should find at least
them knowing many fine things. And in this I was not mistaken, and
they knew things I did not know, and in that respect they were wiser
than me. But, Athenians, they stuck me, the good craftsmen too, as
making just the same error as the poets: because of their fine craft each
claimed to be very wise in others and the greatest of matters too; and
this inconcinnity of theirs hides their actual wisdom; the result was that
I asked myself on the Oracle’s behalf whether I would accept my being
just the way I am, in no respect wise in their kind of wisdom, nor
ignorant with their ignorance, or my having both their qualities. The
answer I gave myself and the Oracle was that I was better off just the
way I am. (Apol. 22c9-e5)

Socrates acknowledges that the craftsmen—unlike the politician and the
poets—have knowledge. The craftsmen are presumably able to give a coherent
explanation of that which falls within their field of expertise, whereas neither
the politician, as I argued above, nor the poets were able to do that. When
Socrates investigates the poets he explicitly bemoans the fact that ‘virtually
almost all the bystanders would have spoken better than them on the subjects
which they themselves had written.’ (Apol. 22b6-8). The poets cannot ‘speak
well,’ because they cannot give a coherent explanation of the content of their
poetry (as opposed to deliver flowery speeches clothed in beautiful language
and embodying fine stylistic juxtapositions). On this evidence Socrates
concludes that the poets ‘say many fine things, but know nothing of what they
say.’ (Apol. 22c2-3).

The poets and the politician lack knowledge – and in so far as te/xnh is
knowledge they also lack te/xnh. The craftsmen are presumably said to know
precisely because they genuinely possess technai which makes them able to
give coherent explanations of their expertise. The craftsmen are also said to be
wiser than Socrates in proportion to the knowledge and te/xnh which they
possess and which Socrates lacks (Apol. 22d3-4), and Socrates asks himself
who is better off – the craftsmen with their kind of wisdom or Socrates (Apol.
22e2-3)? Technê, it seems, amounts to wisdom. However, something peculiar
is going on. Socrates tells us that the craftsmen’s shortfall of self-knowledge
with respect to the greatest matters hides their wisdom – ekeinên tên sofian
apokrupteîn. What does that mean?

Socrates’ attribution of a wisdom which is hidden could be seen to cancel
itself out, and to deny that the craftsmen in fact possess any wisdom, in the
same way in which Socrates’ attribution of a wisdom which is more than
human could be seen to cancel itself out and to deny that the sophists in fact
possess any wisdom. Socrates might well think that the craftsmen are not wise

48
at all if their wisdom is hidden. However, each of the craftsmen uncontroversibly possesses knowledge and technê. If Socrates nevertheless denies that the craftsmen possess any wisdom he must therefore also deny that knowledge and technê by itself amounts to wisdom. The metaphor of hiding, however, suggests that knowledge and technê could be revealed to be wisdom. Since it is lack of self-knowledge which hides the wisdom of the craftsmen it is presumably self-knowledge which would be able to uncover or reveal the technical knowledge of the craftsman as wisdom.

On this suggested reading Socrates’s inquiry into the Oracle is an epistemological inquiry into the concept of wisdom and what it means for a human being to be wise. Socrates initially takes wisdom to be knowledge and technê, and he takes the wisest man to be the man who allegedly knows the most important technê – the technê of virtue – namely the politician. Socrates, however, finds out – not only that the politician does not possess a technê of virtue - but also the politician does not possess self-knowledge. Socrates subsequently reasons that self-knowledge in itself is a kind of wisdom and that he therefore is wise in a non-technical way. Socrates therefore concludes that he is wiser than the politician because the politician is neither wise in a technical way nor in a non-technical way. Socrates reaches this conclusion by realising that wisdom does not need not to be knowledge or technê. ¹ In his subsequent investigation of the craftsmen Socrates concludes something even stronger. Knowledge and technê without self-knowledge is not wisdom; self-knowledge is a necessary condition for both knowledge and technê to amount to wisdom.

There is another way to read the passage. Why, it might be objected, should hidden wisdom not be real wisdom? For example, we might say that a singer’s poor voice hides his good sense of rhythm, but we would not mean that his sense of rhythm is not real. We would simply mean that his poor voice is more striking than his attention to rhythm. So too Socrates can surely mean that the wisdom the craftsmen enjoy is real wisdom, but, from Socrates perspective, less important than human wisdom. Socrates is merely suggesting that there are different sorts of wisdom and that they vary in value.² He is not trying to make the two much stronger claims 1) that knowledge and technê by themselves are not wisdom, and 2) that self-knowledge is a necessary condition for knowledge and technê to amount to wisdom.

The text is indeterminate and can be read either way, I think. But even if we cautiously decide that the second reading is more plausible, Socrates’s inquiry into the Oracle can still be seen as an epistemological inquiry into the concept of wisdom and what it means for a human being to be wise. Socrates concludes that he is better off than the craftsmen. There are, of course, many ways in which you can be better off than somebody else but Socrates thinks that he is better off than the craftsmen because he is wiser than they are. This is clear from the final conclusion he draws:

¹There is nothing in the Apology to suggest that self-knowledge is a technê, although there is evidence to suggest otherwise in other dialogues. See the introduction.
²I owe this musical example, which picks up on plêmmeleia (Apol. 22d8) to an anonymous commentator.
But the probability is, Athenians, that in truth the god is wise, and that in that oracle he is saying this, that human wisdom is worth little or nothing. He seems, moreover, to be talking about this Socrates, and to be using my name in addition, by way of constituting me an example, as if we were to say: ‘That one of you, mortal men, is wisest who, just like Socrates, has realised that in truth he is worth nothing in respect of wisdom.’ (Apol. 23a5-b4)

Socrates concludes that he is wisest, and this entails that the way in which he thinks that he is better off than the craftsmen is by virtue of wisdom. This conclusion rests on the premise that wisdom – or the wisdom which it is possible to possess as a human being – is principally to be thought of as self-knowledge and not as knowledge or technê. If it had been the other way around the craftsmen would obviously have had a better claim to wisdom than Socrates. By declaring himself to be the wisest, Socrates suggests that the wise man is best thought of within the framework of self-knowledge rather than simply in terms of knowledge and technê, and this is a startling epistemological claim, which redefines the way in which Socrates originally thought of wisdom.

Bibliography

Vlastos, G (1985b) ‘Happiness and Virtue in Socrates' Moral Theory’ *Topoi* 4: 3-22
My conception of metaphysics is what could be called ‘Aristotelian’, as opposed to Kantian. In this paper I will examine what Aristotelian metaphysics amounts to and what is its relationship with contemporary metaphysics. The first thing that should be noted is that we are not so much dealing with the details of Aristotle's metaphysical theory – although these as well are relevant at times – but rather with the method that Aristotle used to pursue metaphysical topics. The most important aspect of the Aristotelian method is that metaphysics lies at its heart, i.e. the metaphysical considerations that Aristotle makes affect all other aspects of his philosophy. The idea that metaphysics is necessary for all other philosophical activities is indeed the key point in my conception of metaphysics as well. Quite generally it could be said that this is the upshot of Aristotelian metaphysics: metaphysics is the first philosophy, the starting point for all our philosophical and scientific projects. In what follows we will see how the idea emerges in his work. Aristotle's key works in this regard are Categories, De Interpretatione, Physics and Metaphysics.

The way Aristotle approaches his topics is evidently very closely tied to the basic features of his metaphysics. This can be seen for example in the very beginning of his Physics (1984a: 184a10-184b14), where Aristotle notes that the best way to reach information about the ‘science of nature’ is to advance from universals to particulars, because universals are easier for us to grasp with the help of our senses.\(^1\) Universals and particulars he introduces in De Interpretatione (1963: 17a38). Whether or not Aristotle is right about the role of universals and particulars in our inquiries about reality, it is clear that his account is based on prior considerations about the governing features of reality, namely that our objects of inquiry include both particulars and universals. Many of these prior considerations are laid out in Categories (Aristotle 1963),

\(^1\)For discussion about Aristotle's method in Physics see Bolton (1991).
which is the precursor of category-theory in modern ontology, discussing notions like ‘substance’ (2a13f), ‘quantity’ (4b20f) and ‘relation’ (6a37f). Notions like these are unavoidable in any scientific or philosophical activities\(^1\) and it seems to be quite uncontroversial that philosophers ought to give some kind of an account of them. The manner by which Aristotle handles them is, however, nothing like how Kant does. Unfortunately, Kant's conception of these notions as a part of us rather than the reality continues to burden contemporary metaphysics. The problem is that when this route is taken, we are conceding to the idea of an unbreachable barrier between us and the reality – an idea which effectively leads to relativism. So, what we are faced with now is to consider how the Aristotelian method might be applied to the modern debate and whether the kind of realism that we see in Aristotle is able to cope with the anti-realist tendencies in metaphysics which emerged after Kant.

Aristotle starts *De Interpretatione* with an observation that might be of interest to us. In the following passage he seems to put forward a version of direct correspondence:

\[
\text{[S]poken sounds are symbols of affections in the soul, and written marks symbols of spoken sounds. And just as written marks are not the same for all men, neither are spoken sounds. But what these are in the first place signs of – affections of the soul – are the same for all; and what these affections are likenesses of – actual things – are also the same. (Aristotle 1963: 16a1.)}
\]

We must not let the rather mystical sounding phrasing ‘affections of the soul’ confuse us. Quite simply, ‘affections of the soul’ are thoughts, or, if you like, propositions, whether or not they have been uttered. So, Aristotle suggests that while these propositions can be uttered in a number of ways, say in different languages, the correspondence relation from ‘affections of the soul’ to the actual things always holds between the same terms.\(^2\) Direct correspondence like this surely has its problems, but I think that Aristotle's account is no less tenable than any of its modern alternatives. It is not our task here to argue for this, nor do we need to look at all the details of Aristotle's account, but we ought to keep this background in mind when we examine the Aristotelian method.

Aristotle seems to be foremost interested in the organisation of actual things, and what he presents in *De Interpretatione* (1963) is the method by which we discuss them and some restrictions that apply, for example, to the introduction of modalities. Actual things, according to Aristotle, include particulars and universals (17a38f). It is clear that in Aristotle's ontology, particulars and universals are mind-independent categories in the world, and we refer to them whenever we make affirmations such as ‘every man is white’

\(^1\)In fact, concepts like these are usually presupposed, at least in scientific contexts.

\(^2\)However, Aristotle (1963: 16a10) notes that not every affection of the soul is true or false. Later (17a8f) he specifies that a statement-making sentence, i.e. a sentence that has a truth-value must contain a verb. Aristotle introduces some other restrictions as well, but the main line of thought is very clear: certain 'affections of the soul' have truth-values and they express propositions.
The Aristotelian Method and Aristotelian Metaphysics

(ibid.). This would be an example of stating something universally of a universal (i.e. ‘man’), as Aristotle puts it. This is, very roughly, the connection between his ontology and our language. The importance of De Interpretatione to us is just this: whenever Aristotle mentions a problem in the terms that he introduces in De Interpretatione, we know that he wants to say something about the actual things in the world. This is especially important if one wants to make any sense of his Physics.

As we noted above, Aristotle starts Physics by reminding us about the universal/particular distinction and suggests that we should approach the problems at hand from universals to particulars (contrary to what Plato suggested). It should be quite uncontentious that Physics is deeply involved in what we would certainly call metaphysics. For instance, one of Aristotle's initial concerns is the number of basic principles that govern different kinds of objects (1984a: 184b15 ff.). He dismisses the possibility of there being only one and concludes that there must be three of them (191a20-21). The fact that Aristotle's predecessors thought that the principal elements could include water, fire, air and earth, should not mislead us, although it might render parts of the discussion obsolete. The importance of this passage lies in the attempt to find common grounds for all (material) existence. The suggested explanations might not be correct, but they are logically sound.

So, already on the opening pages of Aristotle's Physics we are very deeply involved with metaphysical questions of the most fundamental sort. There is an obvious explanation for this metaphysical tendency in Aristotle's discussion of natural science. As Aristotle puts it in his Metaphysics (1984b: 1026a13f), natural science is not the first philosophy. There is something prior, an immovable substance, which has to be examined before natural science, which is concerned with movable things, can be pursued. However, natural science would be the first philosophy without the immovable substance. The motive behind this is of course Aristotle's account of tracking movement into the immovable first mover – a view that might be logically sound, but which perhaps seems problematic in the light of modern physics.

Aristotle's Metaphysics is especially interesting for us because there he considers a number of fundamental questions about the nature of metaphysics as a discipline: what are its tasks, method and basis. For Aristotle, metaphysics is the study of the essence of being, being as it is in itself. This is strongly contrasted with something like the Quinian idea that metaphysics should make a complete list of what there is. Rather, Aristotle is interested in what grounds the existence of different kinds of entities, why are they what they are? Furthermore, as Vasilis Politis notes, we must be careful to correctly appreciate what kind of questions Aristotle considers to be relevant for metaphysics:

In general, we must not confuse questions of the type, (1) ‘Why are there things that are F?’, with questions of the type (2) ‘Why are the things that are F F?’ The basic question in the Metaphysics, ‘What is it for something, anything, to be?’, is associated with questions of type 2, not type 1. (Politis 2004: 4.)
Aristotle's view is that natural science is concerned with material, moveable entities. Mathematics, on the other hand, concerns abstract objects. However, neither of these disciplines can be said to be universal, as they are restricted to certain categories of being. It will then be the task of metaphysics to pursue being \textit{qua} being, to examine what kinds of metaphysical constraints govern different kinds of entities. Aristotle proceeds to investigate what being \textit{qua} being might involve and is convinced that the most important category in this investigation is that of substance (1028a30-35). Of the possible ways of how substance relates to entities, Aristotle notes four: essence, the related universal, genus and substratum (1028b33-35). What follows is a detailed account of these features of being. Perhaps of the greatest interest to the modern reader is Aristotle's account of essence, which is clearly the predecessor of the contemporary essentialist views: ‘The essence of each thing is what it is said to be in virtue of itself’ (1029b13-14). It is through the essences of things, and only them, that we can acquire further knowledge about the reality. To be able to determine, for instance, how many objects there are, we must first know what the essences of the objects in question are. It is no surprise then, that essence is what Aristotle calls ‘the primary being’ (\textit{ousia}) (cf. Politis 2004: ch. 7, Loux 1991). It must be noted here though that Aristotle's account, that of metaphysics as the science of essence, is itself a metaphysical answer to the question about the nature of metaphysics. He does consider other possible answers to the question as well, namely that the primary being is either the particular or the universal (and indeed, in the \textit{Categories}, he proposed a different answer). But even if one disagrees with Aristotle about essences being at the centre of metaphysics (which I do not), his method is still very much worth attention. Furthermore, it should be made clear that there are a number of different ways to understand essences. Aristotle's conception is no doubt what could be called ‘metaphysical’ as opposed to ‘semantic’ essentialism: essences are not analytic; they are ‘what is expressed by a complete account of what it is to be for a certain kind of thing’ (Loux 1991: 75; see also Politis 2004: 16 ff.).

So much about the object of inquiry of the first philosophy. This quick overview hardly does justice to Aristotle, but an exhaustive account of Aristotelian essentialism is not necessary for our purposes. We will now turn to the relationship between Aristotelian metaphysics and other disciplines, most notably natural science. Before the inquiry into the second philosophy, i.e. natural science, can start, we must already have done some work in metaphysics. Nevertheless, the topics discussed in \textit{Physics} are of great importance for Aristotle and it is only because natural science is dependent on some more fundamental principles that we have to focus on metaphysics first. We certainly do not have to agree with Aristotle on the details of these principles, although it seems that much of what he contributed to the discussion about essences and universals still survives in contemporary metaphysics. In terms of the relationship between the first philosophy and special sciences we have the following situation:

There is a science which investigates being as being and the attributes which belong to this in virtue of its own nature. Now this is not the
The Aristotelian Method and Aristotelian Metaphysics

same as any of the so-called special sciences; for none of these others deals generally with being as being. They cut off a part of being and investigate the attributes of this part – this is what mathematical sciences for instance do. Now since we are seeking the first principles and the highest causes, clearly there must be some thing to which these belong in virtue of its own nature. (Aristotle 1984b: 1003a22-28.)

The above passage is perhaps even more accurate now than it was when Aristotle wrote it. Special sciences in Aristotle’s time were certainly fewer and a lot closer to what Aristotle himself was doing than special sciences and philosophy are now. However, it is not that the special sciences would be entirely separate from the first philosophy; rather, they concentrate on parts of being that have been cut off from the complete list of entities. Aristotle’s example is mathematics – certainly a part of the science of being, but only concentrating on a small section of it.

Once the limitations of special sciences are admitted and it is acknowledged that even sciences like physics lack the ability to deal ‘generally with being as being’, we then have the possibility to effectively combine our results in metaphysics and special sciences. But how should this be done? Well, in the lines of the Aristotelian method, we should first focus on the most general principles that govern all being and proceed into the details of these principles, such as particular essences and universal attributes of different kinds of entities. After these ontological matters have been settled, we have the tools to interpret the perceptible reality accordingly, i.e. to make sense of the results that we reach in special sciences.

Note that something very important is being said about the basis of metaphysics itself here as well. The way that Aristotle approaches metaphysical topics is in the form of aporiai, philosophical puzzles.1 While metaphysics is about the question ‘What is being qua being’, it is also about the very nature of this question, the possibility of metaphysics. As Politis (2004: 80) notes, it would be a mistake to think that these questions are genuinely separate in Aristotle. For if they were, this would seem to suggest that one can somehow step outside metaphysics, which is not what Aristotle thinks. The importance of this cannot be stressed excessively: Aristotle sees metaphysics as an unavoidable, primary discipline; the questions about the nature of metaphysics are metaphysical themselves and should be treated accordingly. No other discipline – physics, semantics, or even logic – can accommodate the most fundamental questions about the nature of metaphysics, as this would imply going outside the framework of metaphysics. This has numerous important ramifications, for instance, Aristotle's defence of the principle of non-contradiction (PNC) respects this framework, as it is his claim that PNC is the most secure statement about how things are in the world. In other words, it is not a statement about how we think about things or how we talk about them, that is, it is not a logical principle, but a metaphysical one. The upshot of this is that according to Aristotle, logic is grounded in metaphysics, in the ways that things are in the world. Indeed, Aristotle's line of thought

1See Politis 2004: ch. 3 for an extensive account on aporiai.
suggests that the link that is often taken to be between language or grammar, and logic, is really between reality and our thoughts:\footnote{But see Bolton (1994: 350-351) for an important clarification.}

Aristotle argues [in Metaphysics IV.4] that if PNC were not true of things, then we could not use thoughts and words to signify things, and in general we could not think and speak about things. He concludes that if PNC were not true of things, then thought and language about things would be impossible. (Politis 2004: 135.)

Metaphysics, then, is indeed the first science or the universal science. Yet it is worth emphasising that although metaphysics concerns all that is and is universal in this sense, it does not mean that its goal is to reach complete descriptions about all things. The universality of metaphysics is based on the fundamental nature of it, it examines being \textit{qua} being, the preconditions of all being and the governing principles, such as PNC, which affect \textit{all} being. It is the task of special sciences to complete the descriptions, each in their respective field – metaphysics is the study about the common features that range across all disciplines. The question at hand here concerns ‘the metaphysics of metaphysics’; it is about the nature of the question ‘what is being’. Only after this question has been settled can Aristotle offer his answer to the original question of metaphysics, ‘what is being’. His answer to the latter question is of course that metaphysics is the science of essences. This is the distinction between the Aristotelian method and Aristotelian metaphysics – often we are referring to the previous although we talk about Aristotelian metaphysics. For my purposes this does not have very serious implications, as I happen to agree both with the Aristotelian method and with the particular answer to the question ‘what is being’ that Aristotelian metaphysics proposes.

But it is fair to ask, what kind of a bearing does the method described above have on contemporary metaphysics? And what about the level of detail that modern physics has reached, could it not be said that all that is left to do is perhaps to establish the complete, final theory of physics, which would arguably reach the general level of being \textit{qua} being? I think not. For one thing, I believe that a final theory of any kind is an impossibility. That is not how science – or metaphysics, for that matter – works. In fact, the whole concept of a final \textit{theory} is contradictory. A theory is never final, as it should always be open for revision. I should not need to add that in the history of science we have seen plenty of ‘final’ theories which proved out not to be quite so final. Secondly, even if the best approximation of a final theory in physics would be reached, it would in no way render metaphysics redundant. There are two reasons for this: on one hand metaphysics is necessary for interpreting any results reached in special sciences, as some kind of categorisation of the results is needed. On the other hand, metaphysics is and must also be the starting point of any such theory, because surely a theory that claims the title ‘final’ must deal with being \textit{qua} being on the most general level possible, i.e. on the level of the essences of entities rather than on the level of their observable features.
A more serious problem in any attempt to reconcile Aristotelian metaphysics with contemporary metaphysics is perhaps his idea of the immovable substance. Other details of his ontology and organisation of categories that we might not like can easily be dismissed in favour of something else, but the immovable substance seems to be Aristotle's motivation to pursue these topics in the first place and abandoning it would appear to introduce some problems. Perhaps a quick look into the reasons of why Aristotle postulates the immovable substance will help. Clearly, Aristotle is puzzled by motion and one of his basic principles is that there must be a cause for all motion: ‘Everything that is in motion must be moved by something’ (Aristotle 1984a: 241b34). Now, this is a very problematic assumption and very hard to establish in terms of modern physics. Nevertheless, this assumption combined with the assumption that we cannot have an infinite line of movers, which Aristotle (1984a: 241b34 ff.) argues for in some length, produces the conclusion that there must be an immobile first mover. Perhaps this line of thought seems quite untenable now, but I do not think that we can blame Aristotle, for as far off as his line of thought appears to be, modern physics might not do much better. For consider: how does motion emerge according to modern physics?

Well, surely, all kinds of motion can be tracked to material entities. Our current knowledge of all material entities is based on quantum particles: quarks and leptons. Motion enters the picture via forces which are manifested by certain exchange particles. There are four fundamental forces: nuclear strong force, electromagnetic force, nuclear weak force and gravity. For example, the electromagnetic force is manifested through the exchange of photons. A thorough introduction to quantum motion is not required here, but quite generally, all the fundamental forces are exchange forces, as they are manifested through the exchange of one or more particles. And this of course implies motion. But what exactly is the cause of motion according to this theory? There does not seem to be a very straight-forward answer. If we were to look into the details we would find out that there are some dubious cover-ups in effect here. For instance, the exchange particles are ‘virtual’, as they only exist in the exchange process, and in the case of gravity the exchange particle, called ‘graviton’, has not even been directly observed (and it has a rest mass of zero!).

Curiously, as sophisticated and accurate as our current understanding of motion might be, it is blatantly incapable of answering the question that Aristotle asked: how does motion originate? Modern physics provides a number of interesting observations; in the case of motion originating from the electromagnetic force, the motion occurs because there are electrically charged particles present; in the case of motion originating from gravity, the cause of movement is the presence of a body of matter which attracts other bodies of matter nearby. But these are not explanations; they are descriptive accounts about our perceptible surroundings. As far as physics is concerned, there might very well be an immovable first mover which is the one common cause for all motion. What I am saying is that physics does not even try to answer the kind

---

1See for example C.R. Nave (2006) Hyperphysics for details.
of questions that Aristotle puts forward. And this is as it should be, because natural science is, after all, only the second philosophy. There are at least two reasons why one might be unable to grasp this. Firstly, the Kantian tradition has made us too sceptical about the possibility of ever answering these kinds of questions. Secondly, modern science has a peculiar tendency of avoiding questions – it is not the why that scientists are concerned with, but rather what can be observed and what sort of applications it has. However, I think that there cannot be any doubt as to whether we should ask fundamental questions or not. Answering them is the task of metaphysics.

This sidestep to modern physics demonstrates the gap between metaphysics and the special sciences and should help us to see what motivated Aristotle in his judgement that metaphysics deserves a primary status. His method, based on the aporiai, philosophical puzzles, is revealing in this regard: special sciences do not raise general questions about being as such; instead they presuppose that there are different kinds of things ordered in a certain manner. A scientist makes inductive inferences based on perceptual evidence, but by doing so she relies on the orderly nature of reality, she assumes that by certain methods she can come up with veridical judgements about the world. But a metaphysician starts with an abstract puzzle, not an observation – a metaphysician is puzzled about how the scientist can reach knowledge in the first place, how can we know anything about being qua being? This is one of the key questions of metaphysics, and we have seen Aristotle's solution above; his defence of the principle of non-contradiction is especially important in this regard. So, the type of the questions raised in special sciences and metaphysics are radically different. But this is not strictly a difference in their status in regard to the a priori/a posteriori distinction, as one might think. In fact, it would be a mistake to think either that metaphysics is fully in the realm of a priori knowledge or that special sciences are thoroughly a posteriori. Aristotle seems to think that metaphysics and special sciences are fundamentally linked, for metaphysics is the study about the a priori principles that special sciences presuppose. Furthermore, although metaphysics as a discipline is ‘furthest from the senses’ (Aristotle 1984b: 982a25), it is nevertheless continuous with special sciences, and could not operate exclusively in the realm of a priori knowledge.

We are now in the position to see how the Aristotelian method and Aristotelian metaphysics copes with the contemporary challenges to metaphysical realism. Aristotle's central concern is the relativist challenge to fundamental metaphysical principles, such as the principle of non-contradiction. As we saw above, Aristotle thinks that PNC is indeed a metaphysical principle, not a logical principle. What this means is that PNC is one of the constraints that govern the mind-independent reality. For Aristotle, reality is unitary, yet there are different kinds of entities with different essences in the world. PNC is perhaps the most general constraint for the organisation of these different kinds of entities. Plausibly, PNC rules out certain combinations of properties that an entity might have, for instance, no entity can be both green and red all over at the same time, or solid and liquid. The relativist challenges this essentialist, unitary view of the reality by questioning PNC. The modern

---

1See Tahko (2008) for a discussion about related matters.
roots of the relativist challenge can be found in Kant, but Aristotle was well aware of the possibility of such a challenge (cf. Politis 2004: ch. 6).

Aristotle's defence of PNC against the relativist is, as he puts it, a ‘negative’ one: he demonstrates that the opponent's view is inconsistent (Aristotle 1984b: 1006a12). In fact, he goes on to show that the opponent must be committed to PNC at least in the sense that it is true of our thoughts and language (1008b3-1008b32). This is, of course, not enough as such. What needs to be added is that if PNC is true of our thoughts and language, it is also true about the world. Furthermore, the opponent can challenge PNC by pointing out that it often appears, appears to the senses, that is, that the orderly nature of the world required by PNC is violated. To these concerns Aristotle replies as follows:

[I]f only the sensible exists, there would be nothing if animate things were not; for these would be no faculty of sense. The view that neither the objects of sensation nor the sensations would exist is doubtless true (for they are affections of the perceiver), but that the substrata which cause the sensation should not exist even apart from sensation is impossible. For sensation is surely not the sensation of itself, but there is something beyond the sensation, which must be prior to the sensation; for that which moves is prior in nature to that which is moved, and if they are correlative terms, this is no less the case. (Aristotle 1984b: 1010b30-1011a2.)

This is a very dense passage and it is impossible to analyse it thoroughly here. But, clearly, Aristotle is here advocating a realist, causal theory of perception (cf. Politis 2004: 183). He also adds that in fact we never observe a direct violation of PNC in the senses (1010b34-1011a1). This is a crucial qualification, for Aristotle can now justifiably ask, even if the opponent denies the theory of perception that he proposed: how does the relativist explain the orderliness in the world, that is, the observed validity of PNC, which is experienced and apparently true? We must appreciate the weight of this challenge given the context in which Aristotle raises it, for he has argued in length that metaphysics, the science of being qua being, is first and foremost concerned with this very question. Now, if the relativist is to give any kind of a response to Aristotle's challenge, as he must do if he is to avoid being compared to plants [sic] (1006a15), then he is already involved in metaphysics. This is indeed a master argument, for Aristotle has shown here that the only way for the relativist to be involved in a philosophical discussion of any kind is to accept the Aristotelian method and engage in metaphysics. Regardless of what we might think about his particular answers to some metaphysical questions, the Aristotelian method certainly prevails.

References

An Anthology of Philosophical Studies Volume 2


Part II

Modern
Kant’s Rejection of Right of Revolution

Siyaves Azeri, University of Ottawa, Canada

Kant’s denial of right of revolution has bewildered Kant’s scholars.1 Kant holds a sympathetic attitude towards French, American, and Irish revolutionaries. Yet, he rejects the right of revolution. Hence, his stance entails a tension or a contradiction.

The relation between the transcendental consciousness as the principle of the unity of experience and the realm of experience—the nature—is of particular importance to grasp Kant’s view. According to Kant’s Copernican Revolution, the so-called laws of nature are not to be found in nature. Man, as the manifestation of transcendental subject, does not acquire laws of nature via pure experience. Rather, he dictates such laws into the nature. It is not reason that follows the path of nature; on the contrary, it is nature that conveys to rules and laws that are put by transcendental reason. Kant apparently appeals to a dual notion of subjectivity. To the extent that subjectivity signifies the empirical self it is part of the nature; whereas, the subjectivity that dictates laws upon the nature—and therefore upon the empirical self—represents the transcendental subject.

1Among others Reiss (1956), Beck (1971/1978), Axinn (1971), Nicholson (1976), Grceic (1986), and Hill (2002) are the most inspiring accounts of Kant’s attitude towards the right of revolution and its relation to his moral theory.
Kant considers politics with a similar dual outlook.¹ His notion of natural law is based upon Keplerian-Newtonian models and conceptions of universal, natural laws. History, which is the disclosure of social capacities that are implemented by nature in man, is subject to a set of laws that resemble universal natural laws. It is the nature that did ‘bring forth a Kepler who, in an unexpected way, reduced the eccentric paths of planets to definite laws; and then she brought forth a Newton, who explained those laws by a universal natural Cause’ (Kant 1970, 42). Similarly, humans should deal with the constitution of a perfect civil society that regulates the external relations between states conformable to Law. Viewed in this way, the whole social history is conceptualized as the realization of a hidden perfect plan of nature (Kant 1970, 45-6).

However, in conformity to the rules of transcendentality that were laid in the Critique of Pure Reason, these laws and plans, which are inherent in nature, should ultimately be posed upon nature by transcendental subject. When Kant considers ‘the natural principle of the political order’ he evaluates those political principles that are put forward by the transcendental subject of the political realm in order to provide the political orderliness of the civil society: i.e., the natural principle of the political order signifies the transcendental essence of political Right: ‘the will’s manifestations in the world of phenomena, i.e., human actions, are determined in accordance with natural laws, as is every other natural event’ (Kant 1970, 41).

The dilemma of Kant’s approach to revolution can be reformulated similarly: In accordance with the formula of the universal law, it may be concluded that it is inevitable for Kant to approve the right to revolt. Formally speaking, if revolution actualizes the restoration of the rule of the law it is acceptable.

However, revolution aims the very bases of the law and therefore, it should be necessarily disapproved and the right to revolt has to be rejected. Revolution is an empirical act that is conducted in phenomenal realm. The subjects of the revolution are the empirical selves, whom Kant identifies with citizens.

Kant himself formulates this dilemma as the resolution of the question of the priority of formal principles over material principles of action with regard to pure practical reason (1970, 121-2).

On the other hand, the subject of such an action is not and cannot be the empirical self as the member of species but is the representation of

¹Ignoring the distinction between the transcendental subject and the empirical self causes mistaken reformulations of the assumed tension in Kant’s account of the right of revolution. Reiss (1956), for instance, refers to the idea of independence of the individual in face of the authority and asks how, then, was it possible that Kant ‘explicitly denied any individual the right to rebel against established authority’ (1956, 179). These approaches fail to see that the independent individual in civil constitution is the citizen, as a member of civil society and not the individual, natural man as a member of human species. Sidney Axinn (1971), similarly, ignores this distinction and concludes that Kant is a pessimist, when it comes to the individual and an optimist when the human race is concerned (427-8). However, the aforementioned distinction shows that the constitution of the state is the condition of the realization of freedom, equality, and independence of the individual citizen, whereas lack or removal of such a state results in disappearance of these rational principles. It is on the basis of such a distinction that Kant’s rejection of right of revolution follows necessarily from these principles.
transcendental subject. Only the transcendental subject—and the citizen as the representation of this transcendentality that distances himself from the man in the state of nature—can propose and constitute the law. The empirical self should abide and conform to the law of the transcendental. In the "Question of Enlightenment" Kant maintains that as far as the individual is concerned only a few succeed in freeing themselves from immaturity and become free agents (1970, 54-5). However, the chances of public enlightenment are higher (55). This is in line with Kant’s view that contrasts the progress of human species to individual retreat. Kant considers revolutionary action as propagating the yoke of prejudices and immaturity because progress and change toward maturity or enlightenment comes slowly only. Unless all members of the species acquire their share from ‘maturity’ and ‘enlightenment’ we cannot speak of enlightenment, maturity, and freedom. ‘For enlightenment of this kind, all that is needed is freedom’ (Kant 1970, 55). And freedom is the liberty to use one’s reason and thus to ‘argue.’ The ideal enlightened action is rational criticism while obeying the law (1970, 55).

Acting to form an imperative may be interpreted as acting in accordance to the moral law that is rooted in transcendental Right. Moral and free subject acts ethically only if he follows the order of transcendentality. When it comes to choosing between the material and theoretical principles as the basis of righteous political action the latter should be chosen since only the principles of pure practical reason can provide the necessary principles of rightful action. Kant states:

[As a principle of right, it [the aforementioned maxim] has absolute necessity, whereas the former [material principles of action] is necessary only if the empirical conditions which permit the proposed end to be realized can be assumed to exist. (1970, 122)]

This maxim is the formal principle of morality and posits the end of political action based on the concept of duty.

Kant bases the evaluation of political history on a particular conceptualization of man, which is rooted in the Critique of Pure Reason. Kant states: ‘men neither pursue their aims purely by instinct, as the animals do, nor act in accordance with an integral, prearranged plan like rational cosmopolitans’ (1970, 41). Such an approach presumes man having a bifurcated nature that consists of empirical (bodily) and rational (transcendental) elements. This is a symmetric image of the assumed duality between the citizen and the natural man.

The universal order of nature, according to Kant, is essentially teleological; every element in history serves to fulfill the end. What is this ‘end’? The lawful orderliness of nature follows from this teleological making. Reason functions only if there is law, which means that if there is a teleological order. Kant states, ‘those natural capacities which are directed towards the use of his [human] reason are such that they could be fully developed only in the species but not in the individual’ (1970, 42). Therefore the ‘species’ is the representation of the transcendental essence of man in phenomenal realm. The appeal to this transcendentality is caused by natural finitude of human
individual. The essence of this transcendentality which is explained through the language of natural political law is the pure reason:

Nature has willed that man should produce entirely by his own initiative everything which goes beyond the mechanical ordering of his animal existence, and that he should not partake of any other happiness or perfection than that which he has procured for himself without instinct and by his own reason. (1970, 43)

The law of social order—the contract—follows man’s natural situation where he contradicts the benefits of other men. Kant calls this stage the unsocial sociability of man (1970, 44). Kant’s account of the constitution of the social contract stems from Hobbesian premise of the ‘war of all against all.’ In the *Perpetual Peace* Kant clearly expresses this Hobbesian view, which also confirms the interpretation that for Kant human has a bifurcated essence: ‘A state of peace among men living together is not the same as the state of nature, which is rather a state of war’ (1970, 98). Moreover, the ‘individual’ of which Kant speaks is not a socially determined conscious phenomenon but is the member of the species. Such conceptualization of the individual is the necessary consequence of Kant’s naturalism. According to this naturalism man owns a specific, unalterable essence. Man, by nature, is selfish, anti-social, individualist, and immoral. He only pursues his personal well-being and benefit. However, with the rule of reason man transcends this natural crudeness and puts these negativities in the service of the general well-being of human species.

In the ‘Fifth Proposition’ of the *Universal History* Kant answers the question about the ‘end’ that history is supposed to serve. The formation of the civil society that is in concordance with universal Law is the greatest practical problem of the human race. Due to the bifurcated essence of human being there are antagonistic elements at work—human’s natural selfishness and man’s rational capability to pursue the benefit of the species. Therefore, the formation of civil society requires the resolution of this bifurcation.

The highest task which nature has set for mankind must therefore be that of establishing a society in which freedom under external laws would be combined to the greatest possible extent with irresistible force, in other words of establishing a perfectly just civil constitution. (1970, 45-6)

Kant clarifies the nature of this bifurcation when he states that ‘man is an animal who needs a master’ (1970, 46). Where is this master? ‘Nowhere else but in human species’ (1970, 46). The master does not signify the individual person. Rather, it designates the representation of transcendental subject as the source of the transcendental law. The ruler, the administrator, thus, is the embodiment of this transcendentality. Hence, Kant writes: ‘Yet the highest authority has to be just in itself and yet also a man’ (1970, 46). Justice in itself can be acquired only in the noumenal realm, and it is accessible only to the
pure reason—in this case to the pure political reason—as the faculty of the transcendental subject that designates the noumenal self.

The determination of the constitution of history by transcendental subject becomes clearer when Kant speaks about the \textit{a priori} character of the lawful orderliness of the history and contrasts what he calls ‘philosophical history’ to the actual history that consists of the empirical events:

It would be a misinterpretation of my intention to contend that I meant this idea of a universal history, which to some extent follows from an \textit{a priori} rule, to supersede the task of history proper, that of empirical composition. My idea is only a notion of what a philosophical mind, well acquainted with history, might be able to attempt from a different angle. (1970, 53)

Without such philosophical understanding history will be mere historiography that amounts to an arbitrary accumulation of events, which lacks order. The main idea that is inherent in contrasting the philosophical and the empirical histories is that Right, including moral, ethical and political laws, belong to the subject as the owner of the faculty of pure reason.

In Kant’s view, civil society is different from all other forms of contract \textit{not} due to its ends but because of its principles. Civil society is distinguished by being an end in itself; it is not demanded in view of other ends but for the sake of itself that will give way to the establishment of commonwealth. Moreover, the civil constitution of society is presented framed within a particular form of state. The state or the commonwealth is the realization of the rights of man that are secured against the encroachments of others. In this sense, the form of realization of rights of men is ‘external.’

The idea of externality of rights is based upon and reproduces the idea of essential individuality of the form of human existence. Kant addresses this peculiarity of human existence as the phenomenon of ‘self-dependence’ or independence. The only possible relation among men is the mechanical arrangements that regulate the relations among externally posited, isolated atoms. These external regulations have to always be of transcendental nature. Otherwise, they lose the ground of being applicable to the entirety of these monadic individuals.

These regulating laws are limitation of freedom of individuals to which every person subjects himself complying with the authority of the pure reason. Kant, in this respect, states that

Since every restriction of freedom through the arbitrary will of another party is termed coercion, it follows that a civil constitution is a relationship among free men who are subject to coercive laws, while they retain their freedom within the general union with their fellows. Such is the requirement of pure reason, which legislates a priori regardless of all empirical ends (which can all be summed up under the general heading of happiness). (1970, 73)
In the ‘First Addendum’ to *Perpetual Peace*, Kant explains the relation between morality and politics that clarifies the aforementioned distinction between material and theoretical principles. The principles of politics are dictated to men by nature. However, the imposition of political principles does not mean that nature determines human duty. Rather, it signifies the realization of a constitution that teleologically exists in nature. The imposition of duty is realizable only by pure practical reason that acts ‘without any external constraints’ (Kant 1970, 112). Nature functions in favor of man, to comport him out of the state of nature, only practically, since man, by himself, is incapable in this realm (1970, 112).

Moreover, the constitution of state does not occur from the principles of morality. Principles of morality are basically the imperatives that conduct the individual moral actions and are put forward by pure practical reason. These imperatives are assumed to regulate the actions of the individual in relation to his transcendentality. The actions of the individual within a constituted state, on the other hand, are regulated by the transcendental will which is the expression of the Right in the form of law. Principles of morality function internally in contradistinction to the principles of politics that function externally.

The pure, law-giving reason belongs to the transcendental subject. The ‘natural’ universal laws that determine society, like any other natural law, are defined and dictated upon their subjects by the transcendental subject as the unifying principle that also provides the unity of experience. In accordance with this transcendental subjectivism, thus, the particular form of the actualization of transcendental laws that regulate the externally established relation among human individuals, i.e., the particular form of the constitution of civil society, is necessarily monarchy. The monarch, therefore, is the embodiment of transcendental Right that is posited external to other subjects while makes the universal application of laws of society possible. The monarch is the master that is both transcendentally external and exists within the human species.

Monarchy, according to Kant, in contrast to democracy, is farthest from despotism (1970, 101-2). Democracy, meaning the rule of the people by the people, does not realize the law of division of powers.¹ In a democracy, by definition, the same subject, namely the people, actualizes the executive and the legislative powers. Democracy, in its literal sense, cannot assume a

¹Kant distinguishes ‘form of sovereignty’ and ‘form of government’ as ‘forms of state.’ The former can be autocratic, aristocratic, and democratic. The latter is either republican or despotic. Kant, not only does not argue against the compatibility of republicanism and aristocracy, but also devalues democracy as farthest form of sovereignty from the republican form of government. Joseph Grce (1986) seems to ignore this distinction and thus incorrectly identifies republicanism with democracy as the expression of the united will of the people (449). His whole argument about the alleged inconsistency in Kant’s rejection of the right of revolution is based on the assumption that since in Kant’s republican state ‘the legislative power is sovereign and belongs to the will of the people’ then Kant should not hold the view that obeying the law is absolute (450). However, the aforementioned distinction leaves Grce’s inference baseless. Sidney Axinn (1971) also ignores the distinction between the forms of sovereignty and forms of government and supposes a contrast between monarchy and republicanism (426).
representative form. Apparently, Kant’s model of representation is another reflection of his idea about the severance of pure reason and praxis; i.e., Kant’s insistence about the preferability of a representative republican form of government reflects his idea that differentiates between the transcendental subject and the empirical self. Only a representative form of government can guarantee the regulative and external presence of the transcendental will. ‘Act of general will’ determines the form of government (1970, 101) and constitutes men as people. ‘General will’ is not a reflection of an achieved general consent among citizens but is the representation of transcendental will. This act of will is external to every individual man; it is the transcendental principle of the constitution of the people. Therefore, a form of government which is not representative ‘is essentially an anomaly’ (Kant 1970, 101) because it lacks the external, constitutive transcendental will:

We can therefore say that the smaller the number of ruling persons in a state and the greater their powers of representation, the more the constitution will approximate to its republican potentiality, which it may hope to realize eventually by gradual reforms. For this reason, it is more difficult in an aristocracy than in a monarchy to reach this one and only perfectly lawful kind of constitution, while it is possible in democracy only by means of violent revolution. (1970, 101)

Kant enumerates three rational principles of the civil state as liberty, equality, and independence. Liberty is the acceptance of the external limitations upon the individual freedoms that are inherited from the natural state of man for the sake of the formation of the commonwealth. Equality, however, is a totally formal regulation that is determined with respect to the formality of the transcendental law. Equality is the admission of the fact by individuals that they are subject to the coercions that follow the application of Law. Although formally equal, therefore, one can (willfully) be subject to the will of another. This aspect is clearly visible in a monarchy, since to be the subject to the law means being subject to the will of the monarch. Moreover, Kant sees no contradiction between the empirical determination of one subject by the will of another and the formal equality of all subjects before the Law. He states:

This uniform equality of human beings as subjects of a state is, however, perfectly consistent with the utmost inequality of the mass in the degree of its possessions, whether these take the form of physical or mental superiority over others, or of fortuitous external property and of particular rights (of which there may be many) with respect to others. Thus the welfare of the one depends very much on the will of the other… Nevertheless, they are all equal as subjects before the law … as the pronouncement of the general will. (1970, 75)

Independence is the condition of the realization of liberty and equality (Kant 1970, 77). The monadic individual or the citizen—as the representation of the transcendental subject—is the basis of the constitution of civil society.
Due to transcendental subjectivist nature of Kant’s theory, the independence of the transcendental subject appears realizable only in the person of the monarch as the actualization of the law that constitutes civil society. Kant states that it is only himself that the individual does not harm. The individual is also introduced as the master of his own (*sui juris*). However, since the independent individual is the principle of the realization of liberty and equality it is identifiable with the actuality of civil society which is the realization of civil contract. The constitution of civil contract acquires its actuality in the state and therefore in the person of the monarch. Hence, it follows that the fulfillment of the idea of independence of citizens is possible truly within monarch as the representation of the transcendental subject. The head of the state is the representation or the embodiment of pure transcendental subject. Therefore, in case that for instance there is a contradiction between the interests of the citizens and the monarch, since there is no other authority than the monarch by referring to whom the conflict is to be decided, the very conflict should be dismissed (Kant 1970, 81-2). Consequently, the only desirable form of administration of the civil constitution is monarchy.

Principles of politics externally relate and regulate a particular group of empirically existing persons as people. These principles are the universal mechanical laws of the civil constitution. Their ‘naturalness’ does not indicate their non-transcendentality. Rather, it signifies their externality in relation to their objects. In other words, through these principles the naturally given individual is promoted to the level of a member of a civil constitution; he becomes a citizen. To the extent that the citizen conforms to the political principles he will be able to act in concordance with the principles of morality that are dictated by transcendental subject. In this way, the apparent contradiction between the political and moral action is resolved. Political principles make the realization of moral principles possible and not vice-versa. Therefore, the latter does not—and as far as the right to revolt is concerned cannot—contradict the former. Hence, the moral imperatives do not form a basis upon which the right of revolt or rebellion emerges. Moreover, principles of morality are based in the realm of pure practical reason and belong to the transcendental subject. Therefore, as far as their individual application is concerned they presume the existence of the *citizen*, which in turn presupposes the existence of civil society and lawful order. As Kant states.

In their external relations, they [citizens] have already approached what the idea of right prescribes, although the reason for this is certainly not their internal moral attitudes. In the same way, we can expect their moral attitudes to reproduce a good political constitution; on the contrary, it is only through the latter that the people can be expected to attain a good level of moral culture. (1970, 113)

---

1 Thomas Hill (2002) argues against identification of the illegal and immoral in Kant’s political philosophy. He claims that Kant’s basic moral principles might support a position that approves revolution (294). This argument, however, fails to distinguish between man as is considered in his state of nature and man as citizen.
Thus, the right to revolt should be ruled out. Such a right has no theoretical-transcendental justification. Revolutionary action attacks the fundamentals of the transcendental law; it is therefore irrational and destroys the basis of morality; consequently it annihilates freedom. In this sense it is self-deficient.

The idea of social contract, according to Kant, belongs to reason; i.e., the idea of social contract, the constitution of the civil state is established transcendentality. The transcendental essence of the idea of civil society becomes manifest when Kant indicates the impossibility of achieving a full consent concerning the law within empirical realm among citizens. Thus, the relation of the empirically existing citizen to the law is that of duty more than being of consent.

If it is at least possible that a people could agree to it, it is our duty to consider the law as just, even if the people is at present in such a position or attitude of mind that it would probably refuse its consent if it were consulted. (1970, 79)

Thus, the right to revolt should necessarily be rejected. If such a “right” is accepted, then the determination of the transcendental by the empirical is intrinsically confirmed:

Thus if a people, under some existing legislation, were asked to make a judgment which in all probability would prejudice its happiness, what should it do? Should the people not oppose the measure? The only possible answer is that they can do nothing but obey. For we are not concerned here with any happiness which the subject may expect to derive from the institutions or administration of the commonwealth, but primarily with the rights which would thereby be secured for everyone. (Kant 1970, 80)

Man as such and his idea of happiness belongs to the empirical realm. Thus, happiness cannot be referred to “as a principle of legislation”. Principle of happiness signifies the experientiality of the empirical self in its relation to the law. Any action in its name violates the determination of the right of reason which is the law that is dictated by the transcendental subject (Kant 1970, 83). Whereas, men or people must live in commonwealth, i.e., they have to live

---

1Grčič, too, emphasizes the rational-formal nature of the contract: ‘The contract is a formal principle for the evaluation of the justice and lawfulness of the public constitution which has specific material components’ (1986, 448). Yet, his ignoring the distinction between subject of the principle of social contract—that the contract belongs to the realm of transcendental subject—and the empirically existing individual prevents him from acquiring the significance of this formality. He also ignores that Kant defines motives of the will not as ‘objects of physical feelings’ but as ‘absolute law’ (Kant 1970, 68). Grčič, thus, concludes that Kant’s ‘autonomy of will’ is incompatible with his demand for absolute obedience towards the state and that Kant’s demand is justified via a utilitarian argument. He fails to see that for Kant the subject’s absolute duty towards the state follows from state’s being the embodiment of the absolute law. Moreover, he makes the same mistake as Grave and others that Kant criticizes in _Theory and Practice_, who consider the moral duty and the actual deed irreconcilable based on the claim of moral impurity of human nature (Kant 1970, 69).
under the rule of the law that is practically imposed by the state. Every man has the duty to conform to the law of the state that Kant refers to as a transcendental, extra-historical end in itself. Thus, it is the conformity to the transcendental right and not the consent of common citizens as subjects of the rule of the state that legitimizes the law. Therefore, the consent is reasonable only if it is in concordance with the concept of transcendental right.

For so long as it is not self-contradictory to say that an entire people could agree to such a law, however painful it might seem, the law is in harmony with right. But if a public law is beyond reproach (i.e. irreprehensible) with respect to right, it carries with it the authority to coerce those to whom it applies, and conversely, it forbids them to resist the will of the legislator by violent means. (1970, 80-1)

In case that the consent and the law conflict the empirically formed consent of the citizens is to be coerced and outlawed.

It thus follows that all resistance against the supreme legislative power, all incitement of the subjects to violent expressions of discontent, all defiance which breaks out into rebellion is the greatest and most punishable crime in the commonwealth, for it destroys its very foundations. (Kant 1970, 81)

In the Appendix to *Perpetual Peace*, Kant clarifies the relation between political philosophy and moral philosophy: the former cannot contradict the latter and the difference between the two is due to the nature of their objects. Moral principle is the *theoretical* science of right whereas politics deals with its *practice*. Moral philosophy is the science of the pure practical reason. Yet, if the path that is shown by pure reason is followed, the final end of political principle will be clearly visible and the fact that politics and morals are not contradictory becomes evident (Kant 1970, 116). Kant also distinguishes between ethics as the practical science of Rights and morality as the pure theoretical science of Right. He maintains that politics should follow the principles of morality as a pure science (Kant 1970, 129).

Once made, the law should be obeyed and cannot be violated or rejected:¹ ‘the people, under an existing civil constitution, has no longer any right to judge how the constitution should be administered’ (Kant 1970, 81). The assumed right to revolt is an act against the constitution upon which every particular law is based (Kant 1970, 84). Kant upholds that the law is and has to be based upon the principles of pure reason and that, in contrast to the lawful universality of the right, revolution, as an empirical activity, should be rejected. The political right should be ‘based on *a priori* principles, for experience cannot provide knowledge of what is right, and there is a *theory* of political right to which practice must conform before it can be valid’ (Kant 1970, 86).

---

¹Axinn emphasizes this aspect of Kant’s consideration of the state as a form of ‘Victorianism’ (1971, 429).
The idea of inalterability of the right and the state resonates in *Perpetual Peace*. Although Kant does not address the question about revolution directly, his suggestion concerning right of states to exist and not being dissolved into greater states, being taken in its particular historical context, makes his rejection of the right to revolt more conceivable. According to Kant the state is not a thing, neither a property, nor a patrimony (1970, 94). The state is an end in itself and it owns natural rules. Therefore, the limits of the existing states should be left intact. Hence, in the fifth proposition about perpetual peace Kant states that ‘no state shall forcibly interfere in the constitution and government of another state’ (1970, 96).

Ascription of a ‘natural’ essence to state also results in personification of the state. This supports the idea that for Kant monarchy is the most favorable and the ‘natural’ form of administration.

Kant introduces force as an indispensable element of civil constitution, which is caused by the implications of the practicality of politics. When faces inadequacies in the law, a moral politician intends to reform the law in order to attain the ideal political end. Kant explains political inadequacies as practical problems that are caused by *application of the law*; these shortcomings have no roots in the concept of Right. The gradual reform policy that Kant suggests is also a practical means in order to avoid the evils that may be caused by revolution. To the extent that Kant deals exclusively with the practicality of political realm he admits the factuality of revolution. He even maintains that in case that the revolution that is rooted in discrepancies of the older constitution becomes successful and brings along a constitution that is closer to the idea of Right, no attempt should be made to restore the old regime or to punish the revolutionaries (Kant 1970, 118).

Kant does not consider revolution a right but a mere empirical fact. So, the tension between Kant’s sympathetic stance regarding French Revolution and his clear rejection of the right to revolt is resolved. French Revolution was successful as a *matter of fact*. Moreover, the laws of the older regime were not

---

1 Axinn formulates the factual desirability of the revolution as follows: ‘If there is no civil society, then there is no civil law and we may use violence to establish it’ (1971, 426). Contradiction between the government and the law is a form of *contradictio in adjectio*. From a contradiction everything follows, even a revolution. Thus, while Kant does not assume the right of revolution, which means that he does not have a moral-political theory of revolution, he has an analytical phenomenology of revolutions.

2 Thomas Hill (2002) formulates the tension between Kant’s moral principles and matter-of-factness of revolution as the tension between what is morally permissible and what is morally desirable. According to Hill this reflects a more fundamental tension between the inalterability of the Right and historical progress (284). He also maintains that due to this distinction, Kant not only may approve revolutions *ipso facto* but also has a theory that permits revolutions. Peter Nicholson (1976), too, makes a similar point when considering the tension between inevitability of progress and stability of the state, where the latter might block the way of the former. He claims that the question of the inevitability of progress stays unresolved in Kant’s political philosophy. However, the idea of stability of the Right (and of the state) and historical progress are not mutually exclusive in Kant’s view. To the contrary, the factual, historical progress, for Kant, is an approximation to the idea of Right as an idea of Reason. Historical progress is meaningful to the extent that it actualizes the teleological plan of the Reason that functions as the political natural law. Thus, the idea of a Kantian theoretical justification for the right of revolution seems implausible. Yet, factually speaking, to the extent that a revolution
righteous, whereas the post-revolutionary constitution intended to constitute the civil society in concordance with the idea of Right.

The realm of law acquires its unity by the unity of transcendental consciousness. Even in its empirical dimension, the social unity of law depends on the unity of transcendental subject. Revolutionary action reverses the relation between the transcendental subject—as the principle of the unity of experience—and the unity of the experience. It intends to determine the transcendental via the empirical. This is rationally and morally unfounded and thus, it should be rejected.

Moreover, the unity of the self as an element of the unity of the experience also depends on the unity of transcendental subject. Therefore, an empirical act, which contradicts the determination of transcendental subject, amounts to annihilation of the very empirical self, since in such a case the empirical self loses its basis of being.

References


succeeds in constitution of a state that approximates the idea of rational Right it can be approved.
Spinoza on Hatred and the Impotence of Reason

Keith Green, East Tennessee State University, USA

Spinoza provides us with one of the two extensive philosophical "theories" of hatred in western thought before Nietzsche and Freud. I propose reexamining it as a means to better understanding philosophical questions about the nature of hatred as a passion, and its relation to the broader phenomenon of hate. Spinoza is perhaps the first thinker to grasp this relation, and his view is worth reexamining for that reason alone. I will proceed by contrasting Spinoza’s ontology of hatred as a passion, and the sense in which it poses a challenge to the moral life, with that of Aquinas—the one other more extensive pre-freudian “theory”. I will then comment in turn on the three noted contrasts between the ethical upshot of their views. Neither view implies that hatred is what some contemporary philosophers of mind and theorists of emotion call a ‘basic’ emotion. Hatred can be identified as a ‘passion’ in experience only in reference to a ‘folk psychology’, or in reference to an introspectively identifiable set of normative cognitive constituents; and there is less disagreement between Spinoza and Aquinas about the contents of this ‘folk psychology’. Spinoza and Aquinas also agree that persons must never be hatred, though confront us with rival senses of the wrongfulness and irrationality of this hatred.

The Ontology of Hatred: Spinoza vs. Aquinas

Spinoza understands hatred, love, and all affect and emotions, in terms of an ontology, according to which anything’s ‘essence’ includes what one is tempted it call its ‘mode of function.’ This function is reflected in ‘preserving’ itself, and in maintaining and potentiating into the indefinite future, insofar as it can insure its own survival and extend its power causally through acting. This
‘mode of functioning’ Spinoza calls conatus (typically translated by White and Stirling as ‘endeavor’). And it refers, for each thing, to its ‘essence’, that is, to a broad range of capabilities, possessing and exercising which is necessary (and cumulatively sufficient) for a thing’s retaining its identity over time. There is a sense in which, for Spinoza, ‘essence’ can be understood as ‘proper function, so long as we keep in mind that Spinoza does not evidently conceive of ‘essence’ as properties necessary and sufficient for a thing’s instantiating an instance of a ‘general’ and/or ‘natural’ ‘kind’.1 Such ‘proper functioning’, insofar as it may be called ‘activity’, is not ‘intentional’ as such, but for humans many, though not for all ‘essential’ activities, it will entail mental ‘activity’ including consciousness, and so ultimately intentional belief and ‘affect’ states.2 Spinoza’s own claim is encapsulated in Ethics III/Prop. 7, and its demonstration: The effort by which each thing endeavors to persevere in its own being is nothing but the actual essence of the thing itself.

From the given essence of anything, certain things necessarily follow (Ethics I/Prop. 34), nor are things able to do anything else than what necessarily follows from their determinant nature (Ethics I/Prop. 29). Therefore, the power of a thing, or the effort by means of which it does or endeavors to do anything, either by itself or with others—that is to say (Ethics III/Prop. 6) the power or effort by which it endeavors to persevere in its being (conatus)—is nothing but the given or actual essence of the thing itself.

Jerome Neu captures the general point well in a contemporary philosophical idiom (Neu 1977, p71) so long as we remember not to understand ‘effort’ in as an ‘intentional’, mental state. “(W)hat we have before us does not constitute a distinct individual unless it does exhibit a conatus, an effort to maintain itself as a coherent unit.”

Hatred for Spinoza is, as are all ‘affects’, a mental state concomitant with a bodily state, where the subject is conscious of its ‘power’ being diminished by causes outside itself, but having only an unclear, confused idea of those causes. Basic ‘affects’ are different forms of the subject’s consciousness of the effects of objects outside itself, on its body and ‘conatus’, sometimes in the form of

1We should see a reflection here to the earlier Aristotelian notion of ‘ergon’ or ‘proper functioning’ which, for Aristotle, is a kind of embodiment of a ‘form’ or ‘essence’. For Aristotle ‘ergon’ is shared by other particulars in virtue of which they each instantiate and ‘token’ a ‘natural kind’. Spinoza, however, does not speak of one ‘essence’ shared by a number of particulars in virtue of which they instantiate a ‘natural kind’. In Ethics II/Def. 2, he famously—and ambiguously—remarks that “to the essence of anything pertains that, which being given, the thing itself is necessarily posited, and, being taken away, the thing is necessarily taken; or, in other words, that without which the thing can neither be, nor be conceived, and which in its turn cannot be nor be conceived without the thing.” (Gutman ed. p79). Apart from all the pitfalls and implications contained in this definition, the point to make here is that even where (in contrast to Aristotle) ‘essence’ is ‘individual’, it also names for Spinoza the vital functioning of a thing as an organism that must function in a certain way in order to preserve its existence and flourishing.

2By ‘intentional’ here, I mean ‘about’ an ‘object’. Thus, the thing a thought, belief, desire, or feeling responses is ‘about’ is, the technical philosophical sense, its intentional object.
appetites, together with ‘inadequate ideas’ of their causes. In Ethics III/Def. 3, He says, “By emotion I understand the modifications of the body by which the power of acting of the body itself is increased, diminished, helped, or hindered, together with the idea of these modifications”. When this consciousness is of the mind/body’s power being extended (or “passing to a greater perfection”), ‘affect’ is ‘laetitiae’ (joy). But consciousness of its power being diminished, conjoined to an ‘inadequate idea’ of its cause, is ‘tristitiae’ (sorrow). “Joy” is, nonetheless, a ‘passive’ state of the mind, even though it is consciousness of the body’s “passing to a greater perfection”, because in by far the most cases, its cause comes from the outside, and not through the subject’s own internal power of reasoning.\(^1\) In Ethics III/Prop. 9, Spinoza says that ‘besides joy, sorrow, and desire’, “I know of no other primary emotion”.

We also have to consider why and how he subsumes desire under ‘affect’. Desire is consciousness of the body’s being moved causally by objects outside it to possess and consume (or ‘enjoy’) these objects. When the movement of appetite is “considered under the attribute of thought” it is ‘decree’ of mind, but when “considered under the attribute of extention and is deduced from the laws of motion” it is apprehended by us as (causal) ‘determination’. (The Demonstration of III/Prop. 2.) In Ethics III/Prop. 9 he claims that insofar as Conatus or ‘endeavor’ is a function of mind, it is ‘will’, but of the mind and body conjunctively, is ‘appetite’. And desire is ‘appetite conscious of itself’. He comments:

Further, between appetite and desire there is no difference, except that the term desire is generally applied to men, insofar as they are conscious of their appetite, and may accordingly be thus defined: Desire is appetite with consciousness thereof. It is thus plain from what has been said, that in no case do we strive for, wish for, long for, or desire anything, because we deem it to be good, but on the other hand we deem a thing to be good, because we strive for it, wish for it, or desire it.

‘Appetite’ and ‘will’ are part of ‘conatus’--the mind/body’s endeavor to survive into the indefinite future, and to extend its powers of action. But because subjects form only inadequate ideas of the objects that causally move their appetites, desire itself is not the mind’s ‘decree’ arising from its exercise of its power of reason, but is ‘determination’ operating causally upon body to ‘move’ ‘appetite.’ So (considering III/Props 3 and 9) desire is consciousness of ‘appetite’ ‘moving’ subjects to possess, consume, or enjoy objects, many with which it must interact causally in order to survive or extend its vital power. Though Spinoza regards humans as ‘creatures of desire’, it is important to see

\(^1\)“Emotion” in the White and Sterling translation of Ethics into English has been substituted by Gutman in his edition, for “affect” or affectus. From this point on, I will follow White’s original use, since what we call “emotions” or “emotion-types” such as anger, fear, grief, disgust, or hatred involve “affect”, or awareness of a salient bodily change, conjoined to an “idea”, with the proviso that the ‘idea’ involves, among other things, forming certain beliefs about the cause of the bodily change.
that he does not think desire is a necessary constituent of optimal human function.1

The critical ontological difference between Spinoza and Aquinas, as far as hatred is concerned, comes in Aquinas’s claim that anything’s seeking to avert from or destroy anything that is inimical to their survival and flourishing is already for it to exhibit ‘natural hatred’. And in doing so, it ‘obeys’ the creative will of God (i.e. in ‘obeys’ ‘eternal law’). (ST I-II Ques. 29 Art. 1 responsio)

When, for example, an insect or bird (or any other animal, as but one set of examples) averts from, or seeks to destroy, any present specimen of a predatory species, it is exhibiting ‘natural hatred’ and it is functioning according to ‘eternal law’ in doing so as a means to preserve its life and well-being. (See especially Summa Theologica I-II Quest. 29 Arts. 1 & 2 responsios in his discussion of ‘natural hatred’.)

Hatred is understood here as a passion (the creature’s mode of action being actuated by the causal effects of objects outside it) but not as what we could call emotion. ‘Natural hatred’ in creatures other than humans, and perhaps other animals, is a pre-mental and pre-conscious ‘essential’ ‘drive’ or ‘appetite’; but it is still ‘passive’ insofar as it is causally ‘set in motion’ by an object outside the subject. And everything is functioning according to its (the laws of) nature insofar as it is drawn to that which sustains it and its flourishing, and seeks to avert from that which is potentially destructive of it, or inimical to its flourishing, as a well-functioning token of a natural kind. In this respect, it functions as God created and intends it to function, and thus, it ‘obeys’ ‘eternal’ and even ‘natural’ law.

For Spinoza, however, there can be no ‘hatred’ at any ontological level beyond or outside of mental life. ‘Hatred’ presupposes as a condition other cognitive components, or forming an idea of one’s own body (which, Spinoza says, just is the mind), and of causes and effects acting upon and through that body, including all other extended bodies constituting the material environment in which one’s own body lives, as a totality.2 It is not that forming ideas is a feature of uniquely ‘rational’ human life or organism. Indeed, Spinoza seems to envisage the notion that ‘the mental’ is a concomitant of the ‘conatus’ of at least every living thing.3

Insofar as hatred as a state of consciousness of one’s vitality and capability to function (‘conatus’) being diminished from without, but having only inadequate ideas of its cause, it is a form of ‘impotence’ and ‘suffering’ without in any way contributing to function and vitality. In fact, it betokens impotence and suffering at two levels; it is first an awareness of the body’s

---

1See especially IV/Prop.18 Demonstration: (omitting Spinoza’s own citations): “Desire is the very essence of man, that is to say, the effort by which a man strives to preserve his being.” One should note the Note, however, where Spinoza reiterates the superiority of acting ‘from reason’, or willing on the basis of what one knows redound to one’s well-being.

2For what I still regard as the best succinct discussion of this view of Spinoza’s, and its historical-intellectual context, see Hans Jonas “Spinoza and the theory of organism” in Grene (1973), 258-278.

3There is a risk here of reading Spinoza as attributing motivationally salient intentional mental states to non-rational living creatures, just as Aquinas would be also be misread as attributing them to all creatures in claiming that they exhibit ‘natural hatred’ in averting from that which is inimical to their survival or wellbeing.
powers of survival and flourishing ‘suffering’ diminishing effects from objects outside itself. And the accompanying consciousness of this diminishing (‘sorrow’ or ‘pain’) is a matter of ‘suffering’ (of being limited in only being able to have) only an unclear idea of these causes.¹ ‘Feeling’ hatred is not, therefore, a necessary feature of any creature’s ‘drive’ to survive and extend itself through action, even as a tragic necessity for creatures of finite mental reach and limited physical power. The body cannot be made invulnerable, and no knowledge contributing to preserving it or its vitality is mediated by way of passion. Spinoza believes, ultimately, that having ‘adequate ideas’ of the causes of physical suffering alleviates ‘mental’ suffering, in addition to providing for the creature the best vantage point from which to resist threats to survival and minimize physical suffering. So in contrast, again, to Aquinas—and this is a critical point—there is no necessary ontological role for hatred to play in the vital, natural functioning of humans as organisms.

Hatred and Reason

It is significant in Spinoza’s view, that if we knew, or could know, what would redound to our survival and flourishing, even without desiring it, our mind, informed by ‘clear ideas’ and correct reason, would ‘decree’ that we possess or consume what we need anyway, apart from any consciousness of its effects on the body, by way of consciousness of appetite. So any genuine benefit fulfilling desire has is, from the God’s eye point of view, is purely incidental. Yet, to the degree that desire is unrequited, it becomes a source of sorrow, of consciousness of the body’s ‘lack’, together with an inadequate idea of the cause of this ‘sorrow’. As we will see below, this itself amounts, on Spinoza’s view, to hatred.

Spinoza introduces definitions of love and hatred in the note of III/Prop 13. “Love is nothing but joy accompanied with the idea of an external cause, and hatred is nothing but sorrow with the accompanying idea of an external cause. We see, too, that he who loves a thing necessarily endeavors to keep it before him and to preserve it, and on the other hand, he who hates a thing necessarily endeavors to remove and destroy it.” And this definition of hatred is repeated in III/Def. 7. In III/Prop. 28, he extends the definition. There, he says that we ‘endeavor’ to “remove or destroy everything opposed to (joy), or which we imagine conduces to sorrow.” The inclination to do so is just one manifestation of ‘conatus’, or the mind/body’s effort to preserve its own continued existence indefinitely, and to extend and secure its powers of action.

In the Demonstration of this proposition (28), Spinoza appeals to tendencies of the mind that, as he noted in II/Prop. 17, impel us to “strive as much as possible to perceive (those things which conduces to joy) as present or actually existing”. Conversely, “(i)f we imagine that a thing which we believe causes us sorrow, that is to say (note, III/Prop. 13), which we hate, is

¹See Ethic III/Props. 1 & 59 where ‘suffering’ or ‘being effected’ by objects outside itself is opposed to ‘acting’, that is changes occurring through reasoning logically from other already-adequate ideas. Here ‘ideas’ arise in the mind, conjoined to changes effected in the body, through causal effects of objects outside oneself.
destroyed, we shall rejoice (III/Prop. 20), and therefore we shall endeavor to destroy it, or (III/Prop. 13) to remove it from us, so that we may not perceive it as present.” (The citations contained in this quote are Spinoza’s own, hence not underlined.)

Since, however, the mind’s “effort or power in thinking is equal to and correspondent with the body’s effort or power in acting” (appealing to II/Prop. 7), and this is limited, confusion results from something like a tendency to wishful thinking, or belief formation driven by desire rather than reasoning, and by virtue of the fact that the mind’s formation of ideas (‘imagination’) operates in accord with ‘laws of association’ as well as with ‘laws of reason.’ The mind’s tendency to form ‘inadequate ideas’ of objects which cause effects upon and within the body compounds the subject’s vulnerability to both forms of suffering that we identified above.

In Ethics III/Props. 9-14 and in III/Props. 15-28, Spinoza takes up ways in which memory, resemblance, and something like sympathy factor into ‘imagination’ or the formation of ‘confused ideas’ that are conjoined to primary affects to form hatred. The diversity and confusion of ‘ideas’ conjoined to joy and sorrow form the mechanics of the diversity and instability (inconsistency) of the emotional life. These propositions reflect, and sometimes explicitly refer, back to his account in II/Props 16 and 18, of how resemblance, association, and memory function to cause the imagination to generate confused ideas. The claim there was that when a subject’s body has simultaneously been effected by two or more bodies such that, by way of imagination, it forms its ideas of them, any subsequent time an image of one of them comes to mind, for whatever reason, it will tend to remember the others. So in III/Prop. 16, resemblance also is said to lead to the transference of ‘affect.’ He notes that the mind is ‘adverse’ to ‘imagining’ that which causes it to feel sorrow. But hatred is, nonetheless, extended to remembered images, and to anything that, to the mind, resembles an already hated object. Resemblance also raises the possibility of vacillation and ambivalence, where we see a resemblance between something that causes us sorrow and something we love.

Resemblance and memory, then, have the capacity to compound sorrow and suffering to the degree that they are vehicles for extending hatred to a wider range of potential objects through ‘complicating’ or proliferating ‘inadequate ideas’ without there being any ‘natural’ point or benefit in emotion motivating action. But because we believe that our survival and well-being are made more secure by the destruction of what we hate, we feel joy when the objects of our hatred are destroyed, whether or not their destruction actually redounds in any way to our survival or flourishing. (See III/Props. 20 & 22.) Broadly speaking, the belief or ‘idea’ that any object that causes us sorrow, and which we consequently hate, does well even in feeling joy, we then hate that, or those, who (we believe) cause its joy, and love those who (we believe) cause it sorrow or suffering. (See III/Prop. 24 & 26).

The ‘inadequate ideas’ implicit in hatred are further compounded by the fact that (according to III/Prop. 27) we imagine anything we believe is enough like ourselves in its vulnerability and power to be affected by objects more or less as we are effected by them. And this extends hatred to anything that effects such another with either joy or sorrow. Pity is sorrow over the misery, or
sorrow, of one whom we love or whom we compare with ourselves as vulnerable to other objects as we believe we are vulnerable to them. And so we are led to hate the causes of the misery of those whom we pity. This sympathy-like extension of joy and sorrow leads, according to III/Props. 28-30 to forming a feeling of hatred towards kinds of things, given what we believe to be their consistent and ongoing effects on the sorrow or misery of others enough like ourselves. Finally, (III/Prop. 31) when we observe concurrence between what we love, desire, or hate and those responses of others, our inclination to feel them is strengthened; and (III/Prop. 33) we desire the love of others as well. Where another is beloved to us, this desire is further intensified, and we endeavor to “cause it to be affected with joy attended with the idea of ourselves.” (III/Prop. 34.)

Hatred and Anger

The first critical practical ethical disagreement that these different ontologies of hatred entail concerns the relation of hatred and anger, and a subsequent difference over the propriety of anger. One critical way in which Spinoza’s conception of hatred carries forward the earlier stoic conception, and contrasts to the line of thought extending from Aristotle through Aquinas, becomes clear when we examine the different ways in which anger and hatred are differentiated in these two traditions of thought. Aristotle’s differentiation of hatred from anger in Rhetoric ii.4 is taken up and extended by Aquinas. However, like the stoic view of Seneca (De Ira II.3) or the Cicero (Tusculan Disputations IV.vii.19-ix.22) where hatred is regarded as ‘ira inveterata’, or “lust of punishing the man who is thought to have inflicted an injury”, Spinoza sees anger as a form of hatred. In III/Def. 36, “Anger is the desire by which we are impelled, through hatred, to injure those whom we hate.” And vengeance (III/Def. 37) “is the desire which, springing from mutual hatred, urges us to injure those who, from a similar emotion, have injured us.”

Aquinas, by contrast, (in ST I-II Q46 A1, A6 sed contra and responsio) modifies Aristotle’s claim that both anger and hatred give rise to the desire to cause pain to their intentional objects. But anger seeks it as a means to avenge the wrongdoer, “according to the measure of justice”, (See also ST I-II Q47 A2.) where hatred aims to harm or destroy its object as an end in itself, and may take pleasure in this action, or the contemplation of this action, that becomes its own end. (See also ST I-II Q46 A6 responsio, and ST I-II Q48 A2, adv. 1.) In his discussion of vengeance (in ST II-II Q108 A1 responsio), Aquinas claims that the character of intentions mobilized by this passion in the “mind of the avenger” determines whether he ‘crosses the line’ into hatred: “For if his intention is directed chiefly to the evil of the person on whom he takes vengeance and rests there, then his vengeance is altogether unlawful: because to take pleasure in another’s evil belongs to hatred, which is contrary to the charity whereby we are bound to love all men.” Though anger may be virtuous, it poses the moral risk of “morphing” into hatred because it does not always “observe the rule of reason as to the measure of vengeance.” (ST I-II
Q9 A2 responsio, following Aristotle: Nicomachean Ethics 1149bln3.)\(^1\) But Spinoza appears to endorse Seneca’s claim in De Ira XVI.6 that a good judge condemns wrongful deeds, but does not hate them. It is clear that Seneca also does not believe the judge will feel or in any way be motivated by anger. Spinoza (IV/Prop. 43) says “(s)o also the judge who condemns a guilty man to death, not from hatred or anger, but solely from the love for the public welfare, is led by reason alone.

For Spinoza, anger and vengeance fit within the logical and phenomenological framework of sorrow conjoined with the (inadequate) idea of a cause, together with the “natural” propensity, arising out of ‘conatus’, to overcome the perceived causes of sorrow as a way of securing and extending one’s life and well-being. The difference between the ‘aim’ (whether the motivated desire to destroy or overcome is valued as a means or as its own end and ‘pleasure’) and ‘intensity’ are irrelevant. Spinoza defines anger as “the desire by which we are impelled to, through hatred, to injure those whom we hate.” (III/Defs. 36, 37) And vengeance differs only in virtue of the fact that it is said to spring from mutual hatred and is a response to those who have already injured us.

It is, by contrast to Aquinas, the fact that it is the affect of sorrow conjoined to ‘confused idea’ that is relevant with respect to the achievement of human good or perfection—i.e. the extension of functional life and the power to realize one’s capacity for ‘self-causing’. Any state-of-mind falling within the logic and phenomenology of hatred, as Spinoza describes it, already represents a condition of ‘suffering’ and, thus, a diminishing of ‘conatus’, without there being any benefit to self-preservation or flourishing that can only be mediated by being the subject of emotion.

### Hatred and Virtue

To understand why Spinoza believes hatred can never be good or virtuous, and should never even be felt or acted upon, we must keep in mind two critical features of his ‘perfectionism’. The first is his claim that virtue and genuine power, or the optimal extent of ‘conatus’ amount to the same thing.\(^2\) The second is that sociality and its benefits are a necessary concomitant of human flourishing; and it has intrinsic and not merely instrumental value to humans. It is interesting that in IV/Prop. 45, where he argues flatly that hatred can never be good, the bulk of his discussion in the 2 corollaries and 2\(^{nd}\) note focus on contrasting the ‘wise man’s’ enjoyment of ‘moderate and pleasant’ food and drink, natural beauty, the arts, even sports, and nonderisive laughter and

---

\(^1\)I owe this point to Bowlin (1997), pp38-39.

\(^2\)Recall that, in the language of E. M. Curly, the ‘reality’ and ‘perfection’ of any thing is, ‘the same thing’. (This simply restates Spinoza’s own claim in Ethics II/Def. 6: “Reality and perfection I use as synonymous terms.”) The ethical implications of this ontology claim are most clearly reflected in Ethics IV/Def. 8: “By virtue (virtus) and power I mean the same thing; that is (Ethics III/Prop. 7) “virtue, in so far as it is referred to man, is a man’s nature or essence, in so far as it has the power of effecting what can only be understood by the laws of that nature.”
‘merriment’ to the gloomy delight of the envious at the misfortunes of others. And his claim that ‘No God’ would esteem “tears, signs, fears, ad other things of this kind, which are signs of mental impotence” as ‘any virtue’ sounds so close to his description of what the ‘superstitious’ regard as virtue, that one is even more impressed by his association of hatred and a sort of conventional religious morality.

Even more noteworthy, however, is the implication that what the the superstitious (or conventionally religiously moral) esteem as virtue is really just forms of ‘mental impotence’. The implication is that mental potency, or genuine virtue, excludes any sort of positive regard for these states of mind. And indeed, in IV/Prop. 46, he argues that “the man who lives by the guidance of reason” will strive as far as he can do so not only to keep himself from “being agitated by the emotions of hatred”, he will “strive to keep others from being subject to the same emotions.” For this reason, he does not reciprocate hatred with hatred, but with generosity and love. We will return to a discussion of this proposition, and its note, below.

The ethical goal for Spinoza, in contrast to that goal as Aquinas understands it, is not to ‘better habituate’ ‘natural’ hatred, but to overcome ‘bondage’ to them by acting, insofar as possible, upon causes from within oneself. And one does this only by acting upon the basis of reasoning correctly upon the foundation of more or less warranted beliefs that one already holds. This is most clearly stated in the final section of the Ethics (V/Prop. 20, note).

For (IV/Prop. 5) the power of any emotion is limited by the power of the external cause as compared with our own power. But the power of the mind is limited solely by knowledge, whilst impotence of passion is estimated solely by privation of knowledge or, in other words, by that through which ideas are called inadequate; and it therefore follows that the mind suffers the most whose largest part consists in inadequate ideas, so that it is distinguished by what it suffers rather than by what it does, while, on the contrary, that mind acts the most whose largest part consists of adequate ideas, so that, although it may possess many inadequate ideas as the first, it is nevertheless distinguished by those which belong to human virtue rather than by those which are a sign of human impotence.

For Spinoza, when hatred and actions it motivates appear to result in survival or justly sustain or extend the reach on one’s ‘agency’, the coalescence of (the experience of the passion) and a ‘good’ outcome are always only coincidental, a matter of sheer luck. Hatred, as a passion never actually contributes to human good, to survival and flourishing. For Aquinas, however, such a coalescence is to be taken at face-value as an instance of finite creatures functioning (as ‘agents”) according to eternal and natural law (“good is to be done, evil avoided”). The aversion it is in our nature to feel, as finite creatures with a capacity to reason that is limited in its reach, can and must be habituated such that it is proportioned to genuine threats to survival and flourishing to which we are vulnerable, and thus is necessarily part of what creatures need to survive and flourish.
In IV/Prop. 37, note 1, Spinoza, however, contrasts “true virtue” and impotence in the course of an argument that one who seeks virtue for himself will also desire it for other men.\(^1\)

The difference between true virtue and impotence may, from what has already been said, easily be seen to be this—that true virtue consists in living in accordance with the guidance of reason alone, and that impotence therefore consists in this alone, that a man allows himself to be led by things which are outside himself, and by them to be determined to such actions as the common constitution of external things demands, and not to such as his own nature considered in itself alone demands.

The final sentence of this passage may sound like a counsel from, and defense of egoism, especially in light of claims in IV/Prop. 18, that virtue and happiness consist in preserving one’s own being, and is an end in itself.\(^2\) But by ‘power’ Spinoza does not mean having the means to optimize the desires one happens to have, in the manner envisaged by Hobbes.\(^3\) Instead, one’s power is ‘optimized’ by being liberated from motives to act derived from ‘affect’, and therefore from desire, originating in unclear, inadequate ideas of the causal powers of other objects populating one’s environment—the common constitution of external things’. It will consist in ceasing to be moved by desire or hatred, but by being moved by what one knows one needs in order to extend survival and maximize understanding and knowledge of nature. “(i)t follows that we can never free ourselves from the need of something outside us for the preservation of our being, and that we can never live in such a manner as to have no intercourse with objects which are outside us.” (IV/Prop. 18, Note) And indeed it is reasonable to seek what one needs for survival, health, and vitality.\(^4\)

**Conclusion**

Spinoza and Aquinas leave us with an unresolved disagreement about ‘the good’ that accounts for their different ways of comprehending how hatred, though ‘natural’, challenges the realization of virtue. I merely note it here, and do not even suggest a resolution of it. It is this: Even though both agree that the finitude and limited reach—the impotence—of natural human reason, explains

---

\(^1\)This passage should be read, I would argue, as a clarification of IV/Def. 8, where he already equated virtue and power. He goes on to say there, “virtue, in so far as it is related to man, is the essence itself or nature of the man so far as it has the power of effecting certain things which can be understood through the laws of nature alone.”

\(^2\)The proposed reading here takes exception to the ‘egoist’ reading of these passages long put forward by William Frankena (1975), esp. 96. It supports the view of, for example, Neu, who contrasts Spinoza’s claims about self-preservation to the essentially empirical Hobbesian claim that it is simply what anything does, and is thus a ‘first law of nature’ (See Neu 1977, p71).

\(^3\)Neu, op cit.

\(^4\)“It is the part of the wise man, I say, to refresh and invigorate himself with moderate and pleasant eating and drinking, with sweet scents, and the beauty of green plants, with ornament, with music, with sports, with the theatre, and of all things of this kind which one man can enjoy without hurting another. For the human body is composed of a great number of parts of diverse nature, which constantly need new and varied nourishment. . .” (IV/Prop. 45 Note.)
why humans, are subjects of passional hatred, they disagree about whether (in
the ‘grand design’ of ‘nature’) feeling and acting upon hatred ever actually
engenders the ‘good’—for Spinoza, survival and maintaining or extending the
reach of autonomous agency.

For Spinoza, when hatred and the actions it motivates appear to result in
survival or justly to sustain or extend the reach of one’s ‘agency’, the
coalescence of the experience of hatred and a ‘good’ outcome from the actions
it motivates are always only coincidental, a matter of sheer luck. Hatred, as a
passion never actually contributes to the human good. For Aquinas, however,
such a coalescence is generally to be taken at face-value as an instance of finite
creatures functioning (as ‘agents’) according to eternal and natural law (“good
is to be done, evil avoided”). The aversion it is in our nature to feel, as finite
and rational, but fallen creatures, can and must be habituated so that it is
proportioned to genuine threats to survival and flourishing to which we, as
finite and fallen rational creatures, are vulnerable. Though for the genuinely
virtuous, this natural aversion never takes the form of hating persons. The
disagreement with which Spinoza and Aquinas leave us confronts those within
theistic traditions of religious ethics and beyond. Even if it can be shown that,
say, ‘hatred behavior’ conduces to survival, or that evolution ‘selects’ for it
because it does, is it ‘rational’, ‘justified’ or ‘good’ if, as Spinoza argues, the
capacity to reason enables humans to achieve the same ends by other means?

Works Cited

Thomas Aquinas (1981), Summa Theologica, trans. Fathers of the English Dominican
Province, Westminster, MD, Christian Classics.

John R. Bowlin (1999), Contingency and Fortune in Aquinas’s Ethics, Cambridge
University Press.


reprinted in Mandelbaum and Freeman (1974).

Eugene Freeman and Maurice Mandelbaum, eds., (1975), Spinoza: Essays in
Interpretation, LaSalle, Illinois, Open Court Publishing Company.

Marjorie Grene, ed. (1979) Spinoza, A Collection of Critical Essays, University of
Notre Dame Press.

Paul E. Griffiths (1997), What Emotions Really Are, Chicago, University of Chicago
Press.


and the Relationship of Philosophical Theories of Emotions to Psychological
Theories of Therapy, Berkeley and Los Angeles, University of California Press.

Benedict De Spinoza, New York, Dover Publications.

Baruch De Spinoza (1949) Ethics, Preceded by On the Improvement of the
(1883), revised Amelia Hutchinson Stirling (1894, 1899), New York, Hafner
Publishing Company.
Kant’s standard formulation of the distinction between analytic and synthetic judgment is given in the Introduction to the *Critique of Pure Reason*:

In all judgments in which the relation of a subject to the predicate is thought,… this relation is possible in two different ways. Either the predicate $B$ belongs to the subject $A$ as something that is (covertly) contained in this concept $A$; or $B$ lies entirely outside the concept $A$, though to be sure it stands in connection with it. In the first case I call the judgment **analytic,** in the second **synthetic** (A 6-7/B 10-11).\(^1\)

The analytic-synthetic distinction is described here in terms of whether or not the predicate is contained in the subject concept; if it is, the judgment is analytic, if not, synthetic.

That the analytic-synthetic distinction is crucial for Kant’s transcendental philosophy should be patent. As Lewis White Beck remarks:

The *Critique of Pure Reason* is supposed to answer the question, How are synthetic judgments *a priori* possible? But if it is not possible to decide objectively whether a given judgment is synthetical or analytical, the entire *Critique* seems wasted effort.\(^2\)

---

\(^1\)References to Kant’s writings are to the pagination of *Kants Gesammelte Schriften, herausgegeben von der Deutschen Akademie der Wissenschaften,* 29 vols. (Berlin: de Gruyter, 1902–), cited as ‘Ak’, followed by volume and page number. All quotations from the *Critique of Pure Reason* are from Kant (1997). I follow the standard practice of citing the pagination of the first (1781) or ‘A’ and second (1787) or ‘B’ editions.

\(^2\)Beck (1984b), 301.
Kant’s distinction has been subjected to many forms of criticism. There have been criticisms from the rationalist as well as from the empiricist, pragmatist, and semantic points of view. In fact, some of the major criticisms of Kant’s distinction were introduced as early as seven years after the publication of the first edition of the *Critique of Pure Reason* (1781) in a journal called the *Philosophisches Magazin*, which was founded in 1788 by the Wolffian philosopher and professor at Halle, Johann August Eberhard. In an essay titled, ‘On the Distinction of Judgments into Analytic and Synthetic,’ Eberhard argues that Kant’s distinction is nothing more than a rehashed, and even a confused, version of the old Leibniz-Wolffian distinction between identical and non-identical judgments. A more incisive criticism of Kant’s distinction, however, was introduced by one of Eberhard’s collaborators, J. G. Maass. Maass argues that Kant’s own distinction, as it is formulated in the Introduction to the *Critique of Pure Reason*, is without any objective and determinate basis. According to Allison, Maass formulated objections that anticipate ‘some of the subsequent and now “classical objections” to Kant’s theory.’

Maass’s critique may be summarized as follows: Kant says that in analytic judgments the predicate is, while in synthetic judgments the predicate is not contained in the subject concept. But how are we to determine whether or not a concept is really contained in another concept? How are we to decide which concepts are and which concepts are not already contained in a concept? Kant’s answer is that we must carefully attend to what we ‘actually think’ in a concept, such that what we do not think in a concept is not included in the content of that concept, and what we do ‘actually think’ is included in it. At this point, however, it becomes evident that the distinction between analytic and synthetic judgments is merely relative to what a person thinks or does not think in the concept of the subject. And on this criterion, nothing can dispel the conclusion that one and the same judgment would be regarded as analytic by one person and synthetic by another, depending upon what each person happens to think or not think in the concept of the subject. As a consequence, Maass concludes that, if Kant’s distinction is to have any determinate and reliable basis (rendering it impossible for the same judgment to be both analytic and synthetic), ‘then a universal rule must be provided, whereby it is in each case possible to determine whether or not the predicate B lies in the subject A ….’ Since, however, argues Maass, such a rule is not provided by Kant, his distinction remains ‘completely relative and of no real philosophical importance.’ If the entire *Critique of Pure Reason* is not to be regarded as

---

1Eberhard (1789), 321-22; Ayer (1952), 78; Quine (1953), 21; Coffa (1991), 17.
2Allison (1973), 42.
3Ibid.
5What is perhaps more devastating is that Kant’s criterion cannot avoid the possibility all synthetic judgments are covertly analytic. Leibniz held that in all true propositions, the predicate is included in the concept of the subject. In other words, all true propositions are analytic in the Kantian sense. See Beck (1975), 12.
6Maass (1789), 190, quoted in Allison (1973), 43.
7Allison (1973), 43.
‘wasted effort,’ Maass’s criticism must receive an answer. Did Kant provide an answer?

Although Kant himself did not respond to Maass’s critique, Maass’s essay was examined by Kant’s supporter and friend J. G. Schultz, who wrote his reply ‘under Kant’s supervision.’\(^1\) Lewis White Beck first examined Schultz’s reply to Maass’s critique in two essays, ‘Kant’s Theory of Definition’ (1956) and ‘Can Kant’s Synthetic Judgments be Made Analytic?’ (1955). These essays have provoked several other Kant scholars to engage in a thorough investigation into the nature and problem of Kant’s distinction, which has continued to this day.

It is my aim in this paper to provide an answer to the question, What are the grounds for the analytic-synthetic distinction as Kant himself understood it, and are they adequate to defend Kant against Maass’s objection? I first examine briefly the attempts made by three scholars, Lewis White Beck, Moltke S. Gram, and Henry Allison, to respond to the objection raised by Maass, as well as to demonstrate the validity of Kant’s distinction. In part II, I show that Kant provides another formulation of the distinction, a formulation that has escaped the notice of the scholars in question. In part III, I show that this formulation provides Kant with a ground for proving the irreducibility and objective validity of his distinction, a ground that is based on the distinction between general and transcendental logic.

Lewis White Beck, Moltke S. Gram, and Henry Allison all agree that in Kant’s formulation one can detect two criteria for deciding whether or not in any given judgment a concept is ‘contained’ in another concept: the so-called ‘logical’ and the ‘phenomenological’ criterion.\(^2\) All three scholars agree, moreover, that neither criterion is sufficient to establish the objective validity of Kant’s distinction. They point out, for example, that according to the logical criterion, while all analytic judgments are, synthetic judgments are not justified by the principle of contradiction alone: ‘we must … allow the principle of contradiction to count as the universal and sufficient principle of all analytic cognition’ (A 151/B 191). According to this criterion, to know whether the predicate is contained in the subject concept, we must ask whether negating the judgment leads to a self-contradiction. If it does, the judgment is analytic, if not, synthetic.\(^3\) As the scholars point out, however, the problem with this criterion is that it cannot be applied in the first place without appealing to the other criterion mentioned above, namely, the phenomenological criterion.\(^4\)

According to the phenomenological criterion, to know whether a concept is contained in another we must pay attention to what we ‘actually think’ (B 17) in the concept, such that what we do not think in a given concept is not included in the content of the concept, and what we do ‘actually think’ is included in it.\(^5\) As we have already seen, however, at this point, the distinction between analytic and synthetic judgment becomes merely relative to what a

\(^{1}\) Beck (1984a), 313.
\(^{3}\) Allison (1983), 74.
\(^{4}\) Ibid., 74-5.
\(^{5}\) Beck (1984a), 307-10
person thinks or does not think in the concept of the subject. In other words, Kant’s distinction remains vulnerable to Maass’s criticism.

Convinced of the inadequacy of Kant’s standard formulation of the distinction, all three scholars have looked for other clues in Kant’s writings in order to defend his distinction. I would now like to look briefly at each scholar’s defense of Kant’s distinction, starting with that of Lewis White Beck, whose analysis is developed further by Henry Allison. I close Part I with an examination of Moltke S. Gram’s interpretation.

Lewis White Beck presents his defense of Kant’s distinction via a commentary on the aforementioned reply made by Schultz to Maass’s criticism. In his essay ‘Kant’s Theory of Definition,’ Lewis White Beck draws our attention to Kant’s distinction between logical and real predicates or ‘determinations.’ Beck points out that, according to Kant, an analytic judgment contains ‘a logical predicate,’ which ‘states the logical essence of the concept of the thing, or serves merely to distinguish this thing from others,’ whereas a synthetic judgment contains ‘a real predicate’ or ‘a determination’ (Bestimmung), which ‘determines a thing, and not merely its concept.’

In another essay, Beck offers a more complete defense of Kant’s distinction. He begins by introducing Kant’s distinction between the logical and the real possibility of a concept. A concept is logically possible if it is merely shown that it is not self-contradictory. To be really possible, on the other hand, ‘the concept must be of an object which “agrees with the formal conditions of experience”… and this includes the condition of sensible intuitability.’ According to Beck, an analytic judgment is one in which the subject is a logically possible concept and the predicate expresses a property of the logical essence of such a concept of the subject. A synthetic judgment, on the other hand, contains a concept that is really possible, and therefore, ‘includes the condition of sensible intuitability.’ Thus, what Beck’s account underscores is that the main difference between analytic and synthetic judgment consists in the fact that while an analytic judgment relates one concept with another concept, a synthetic judgment involves a concept that also refers to an intuition (either empirical or pure). Henry Allison also emphasizes this difference. He says that for analytic judgments the question whether a concept refers to an object is ‘otiose’; thus ‘the “reality” of the predicate does not come into consideration in the judgment.’ A synthetic judgment, on the other hand, relates a concept to an object, which, for Kant, means that it is related to an intuition.

In fact, Beck and Allison’s conclusion is in accordance with a remark Kant makes in a letter written in 1789 that ‘all synthetic judgments of theoretical knowledge are only possible through the relation of a given concept to an intuition.’ Does Beck’s (and Allison’s) account, then, succeed in providing

---

3Beck (1975), 21.
4Ibid.
5Allison (1983), 75.
6Ibid., 77.
7Ak, 11: 38, quoted in Allison (1983), 77.
objective grounds for Kant’s analytic-synthetic distinction? According to Moltke S. Gram, the answer is no.

Gram admits that the distinction emphasized by Beck between logical and determining predicates is ‘informative'; however, he also maintains that ‘the information it gives us is not the information we need if we are to distinguish between covertly analytic and genuinely synthetic judgments.'¹ For one thing, according to Gram, to say that a concept in a judgment is a real predicate or determination still does not preclude the possibility that such a concept is included in the concept of the subject. Moreover, Gram contends that though Kant’s distinction between logical and determining predicates may illuminate the nature of the relation between concepts and the objects to which they refer, it does not help to clarify the nature of the relation between two concepts in a judgment. As far as Gram is concerned, what we need to know is not how a concept relates to an object, but how the relation between two concepts can be said to be irreducibly synthetic. Again, as Gram points out, the only criterion Kant provides for determining the synthetic nature of the relation between the two concepts is the phenomenological criterion, which, as Gram shows (and as we have seen), is inadequate.² He concludes, therefore, that not only on Beck’s but even on Kant’s account, the objective validity of the analytic-synthetic distinction cannot be established.

Gram proceeds to develop his own ‘resolution’ to what he calls ‘the crisis of syntheticity.’ But unfortunately his resolution requires ‘an alteration in [Kant’s] theory of judgment,’³ which alteration, I believe, is not only implausible but unnecessary.⁴ Instead of proceeding further, therefore, I would like to close this part of the paper with the following remark: Despite the implausibility of Gram’s ‘resolution,’ the questions he raises are important; for, Gram is right to insist that no account of Kant’s distinction can be regarded as complete that does not identify the very grounds on which the distinction between two (according to Kant, irreducible) ways of relating one concept with another concept is based. I disagree, however, with Gram’s conclusion that the demonstration of the insufficiency of the phenomenological criterion is an admission that the validity of Kant’s distinction cannot be established.

What the scholars have not observed is that Kant has another formulation of the distinction, according to which all analytic judgments, whether theoretical or practical, are species of indeterminate judgments. But in order to understand the significance of this statement, we must first examine Kant’s distinction between general and transcendental logic.

---

¹Gram (1984a), 327.
²Ibid., 324-26.
³Ibid., 343.
⁴Gram proposes that an intuition rather than a concept be made to serve as the subject term of a synthetic judgment. But Thompson (1972) notes that a Kantian empirical intuition is an apprehension of ‘a spatiotemporal something,’ but not of an object determinate enough to be designated by any term in a judgment. He maintains that a Kantian judgment is always ‘a relation between two concepts rather than between a concept and an object subsumed under it’ (p. 325). See also Pippin (1982), 140ff.
In the *Critique of Pure Reason* Kant distinguishes between general and transcendental logic. He says that while general logic ‘concerns the use of the understanding without regard to the difference of objects,’ transcendental logic concerns ‘the laws of the understanding and reason, but only insofar as they are related to objects *a priori*’ (A 57/B 82). On the basis of the distinction between the two logics, Kant explains that the same categorical form of judgment (‘S is P’) can be employed in two different ways, namely, determinately and indeterminately:

[The categories] are concepts of an object in general, by means of which its intuition is regarded as determined with regard to one of the logical functions for judgments. Thus the function of the categorical judgment was that of the relationship of the subject to the predicate, e.g., ‘All bodies are divisible.’ Yet in regard to the merely logical use of the understanding, it would remain undetermined which of these two concepts will be given the function of the subject and which will be given that of the predicate. For one can also say: ‘Something divisible is a body.’ Through the category of substance, however, if I bring the concept of a body under it, it is determined that its empirical intuition in experience must always be considered as subject, never as mere predicate; and likewise with all the other categories (B 128-29).

In this passage Kant explains that there are two ways in which the same logical form of thought can be employed in the act of judging, namely, determinately and indeterminately. It is crucial to point out that in this passage what Kant calls ‘the logical use’ of the categorical form of thought is identified with the indeterminate nature of its function in judging. In yet another passage, Kant, in speaking about a proposition involving the concept of substance, says that ‘the concept of a subject is here taken merely logically, and it remains undetermined whether or not substance is to be understood by it’ (B 419). It may be inferred from these passages that whereas general logic deals with the indeterminate use, transcendental logic deals with the determinate use of the logical form of judgment.¹

The determinate-indeterminate distinction with regard to the employment of the logical form of thought is also elaborated in a footnote in the Preface to the *Metaphysical Foundations of Natural Science*. Kant writes:

The table of the categories completely contains all the pure concepts of the understanding as well as the formal actions of the understanding in judgments, from which such pure concepts are derived and from which they also differ in nothing except that in the concept of the understanding, an object is thought as determined in regard to one or the other function of judgments. (E.g., in the categorical judgment ‘the

¹It is true that Kant also speaks of “determination” in general logic, as, for example, when a lower concept is determined by a higher one in a genus/species relationship (*Ak*, 9: 146). But what is thus determined by or “contained under” a concept is yet another concept. From the standpoint of transcendental logic, in others words, logical determination still leaves undetermined whether the concept relates to an object.
stone is hard,’ the ‘stone’ is employed as subject and ‘hard’ as predicate, so that it remains permissible for the understanding to interchange the logical function of these concepts and say ’something hard is a stone’ [italics mine]. On the other hand, when I represent to myself as determined in the object that the stone in every possible determination of an object, and not of the mere concept, must be thought only as subject and the hardness only as predicate, the very same logical functions now become pure concepts of the understanding for cognizing objects, namely, substance and accident).

Kant explains in this passage that there are two ways in which the same logical form of thought can be employed in the act of judging, namely, in a determinate way, for ‘cognizing objects,’ and in an indeterminate way, for relating concepts in mere ‘thought.’ In other words, whether or not the logical form of judgment applies to an object (i.e., is used as a category) depends on whether or not it is determined which of the two concepts is given the function of the subject and which is given the function of the predicate. The logical use of the understanding leaves it undetermined which concept will be given which function.

It seems evident from the foregoing analysis that the logical forms of thought are always employed in an indeterminate manner in all analytic judgments; since analytic judgments require no intuition, they relate concepts in mere ‘thought’ (not ‘in the object’). For example, to revert to the first passage from the Critique quoted above, Kant says that the judgment ‘all bodies are divisible’ involves the ‘logical use of the understanding,’ which means that ‘it remains undetermined to which of the two concepts the function of the subject, and to which the function of predicate, is to be assigned. For we can also say, ‘Something divisible is a body’’ (B 128-9). In a letter written in 1789, Kant says explicitly that the judgment, ‘every body is divisible,’ is an analytic judgment, for the predicate concept of divisibility is already contained in the subject concept of body. It is clear, therefore, that all analytic judgments are based on ‘the logical use of the understanding,’ which involves the indeterminate employment of the logical function of thought. Further textual evidence for the claim that analytic judgments are indeterminate will be provided below. Suffice it to say for now that while analytic judgments require an indeterminate employment, synthetic a priori judgments require a determine employment of the logical forms, which employment alone makes possible the relation of concepts to an object. In other words, synthetic a priori judgments are species of determinate judgments.

To provide further support for the foregoing interpretation, I would like to examine Kant’s account of the relation between the analytic-synthetic distinction and the determinate-indeterminate distinction in his practical

---

1Kant (1985), 13; Ak, 4: 475n. See also Critique of Pure Reason, A 245-46.
2See also Prolegomena, Ak, 4: 299-300 and Critique of Pure Reason, B 142.
3Longuenesse (1998) says that in an analytic judgment, which has ‘no relation to a sensible intuition,… no category is involved’ (79).
4Kant (1967), 138; Ak, 11: 35.
philosophy. It will be shown that the nature of this relation is the same in both Kant’s theoretical and practical philosophy.

As we have already seen, Kant explains in the Critique of Pure Reason that whereas general logic ‘concerns the use of the understanding without regard to the difference of objects,’ transcendental logic concerns ‘the laws of the understanding and reason, but only insofar as they are related to objects a priori’ (A 57/B 82). In the Groundwork of the Metaphysics of Morals, Kant tells us that while general practical philosophy ‘consider[s] motives without regard to the difference in their source,’ the metaphysics of morals considers a ‘will of a particular kind, such as one determined without any empirical motives by a priori principles; in a word, ... a pure will.’¹ In fact, Kant says that general practical philosophy ‘differs from a metaphysics of morals in the same way that general logic is distinguished from transcendental philosophy.’² In the Ethik Mrongovius³ Kant also remarks, ‘General practical philosophy is related to morals as logic is to metaphysics.’⁴ Kant continues:

In general practical philosophy it is undetermined, whether we have motivating grounds or not. Logic abstracts from our cognitions. General practical philosophy [abstracts] from the grounds of motivation.⁵

It should be noted here that, as this passage from the lectures on ethics indicates, for Kant, to say that a judgment leaves something ‘undetermined’ is equivalent to saying that it ‘abstracts from’ something. In fact, besides the phrase ‘abstract from’ (abstrahir von), there are at least three other phrases Kant uses to express the same notion about the indeterminacy of judgments: ‘without regard to’ (ohne Rücksicht auf), ‘indifferent to’ (gleichgültig), and ‘leave undecided’ (unausgemacht).⁶

To what precisely does the distinction between determinate and indeterminate functions of willing refer? The answer is provided in chapter 2 of the Critique of Practical Reason. Before introducing ‘the table of the categories of freedom,’ Kant explains:

these categories concern only practical reason in general and so proceed in their order from those which are as yet morally undetermined and sensibly conditioned to those which, being sensibly unconditioned, are determined only by the moral law [italics mine].⁷

Kant makes it clear in this passage that the determinate-indeterminate distinction in his practical philosophy represents the distinction between non-moral and moral forms of willing, that is to say, between hypothetical and

---

¹Kant (1990), 6-7; Ak, 4: 390-91.
²Ibid., 7; Ak, 4: 390.
³This lecture was given in the winter semester of 1784-5. See Kant (1997b), xvii.
⁴Lectures on Ethics, Ak, 29: 597.
⁵Ibid., 598.
⁶See ibid., 598; Critique of Pure Reason, A 290/B 346; Ak, 5: 105; and Ak, 4: 324.
⁷Kant (1956), 69; Ak, 5: 66.
categorical imperatives. In other words, general practical philosophy considers hypothetical imperatives, which are those forms of willing that are based on ‘empirical motives,’ or, according to the passage from the *Critique of Practical Reason* quoted above, ‘those which are as yet morally undetermined and sensibly conditioned.’ In the *Ethik* Mrongovius, Kant says that ‘general practical philosophy exhibits the rules whereby the will is determined *a posteriori*.’ On the other hand, those forms of willing that are ‘sensibly unconditioned [and] determined only by the moral law,’ or in a word, the categorical imperatives, are considered by the metaphysics of morals.

It should be evident by now why Kant calls hypothetical imperatives ‘analytic’ practical propositions, and categorical imperatives ‘synthetic *a priori*’ practical propositions. For Kant, all analytic propositions, theoretical as well as practical, leave something undetermined. For example, in chapter 3 of the Analytic of Principles of the *Critique of Pure Reason*, Kant says the following about ‘an analytic assertion’:

> since [an analytic assertion] is occupied only with that which is already thought in the concept, it leaves it undetermined whether the concept even has any relation to objects, or only signifies the unity of thinking in general (which entirely abstracts from the way in which an object might be given); it is enough for him to know what lies in its concept; what the concept might pertain to is indifferent to him (A 258-59/B314-15).

Thus, all judgments are analytic which ‘abstract from,’ ‘leave undetermined,’ ‘are indifferent to,’ or in a word, leave undetermined whether there is any relation to objects. As has been indicated above, these phrases are used interchangeably by Kant. In the *Logic*, moreover, Kant, in distinguishing between analytically and synthetically ‘universal rules,’ says that while analytically universal rules ‘abstract from differences,’ synthetically universal rules ‘attend to distinctions and consequently determine in regard to them too.’

In sum, for Kant, just as theoretical propositions which leave undetermined the question about the relation to an object are analytic, so practical propositions which leave undetermined the question of whether there is a motivating ground of the will are analytic. And just as theoretical propositions that determine an object are synthetic *a priori*, so categorical imperatives are regarded as synthetic *a priori* practical propositions because they involve a determination of the will ‘only by the moral law.’

I would now like to apply the results of the foregoing analysis to the question regarding the grounds for proving the validity of Kant’s analytic-synthetic distinction in his theoretical philosophy. As we have seen in part I, both Beck and Allison’s accounts were vulnerable to the criticism raised by

Gram, namely, that to distinguish (as they do) between a judgment (analytic) which relates a concept with another concept and a judgment (synthetic) which relates a concept with an object does not prove that they are irreducible to one another. Once one takes into consideration the fact that, for Kant, in all analytic judgments the logical forms of thought are employed indeterminately, while in all synthetic a priori judgments the logical forms of thought are employed determinately, it becomes clear why Kant believed that the two kinds of judgment are irreducible to one another. The way in which the predicate and the subject concepts in a judgment relate to one another depends on the way in which the logical forms of thought are employed in the act of judgment. As Kant explains, if the logical form of thought is employed indeterminately, the two concepts in the judgment are related in such a way that it is left ‘undetermined which of these two concepts will be given the function of the subject and which will be given that of the predicate’ (B 128-29); and this indeterminacy with regard to the manner in which the predicate and the subject concepts relate to one another explains why Kant says that one can form an analytic judgment in which ‘anything we please can be made to serve as a logical predicate; the subject can even be predicated of itself’ (A 598/B 626). If the logical form of thought is employed determinately, that is, as a category, on the other hand, the two concepts in the judgment are related in such a way that ‘it is determined that its empirical intuition in experience must always be considered as subject, never as mere predicate’ (B 129). Hence, the synthetic relation between the two concepts in a judgment is made possible by the determinate employment of the logical forms of thought, and this relation is irreducible to the analytic relation between the two concepts in a judgment, a relation that is based on the indeterminate employment of the logical forms of thought. Since the determinate-indeterminate distinction is grounded on the distinction between general and transcendental logic, the validity of Kant’s analytic-synthetic distinction depends ultimately on the validity of the distinction between general and transcendental logic.

Bibliography

Works by Kant

Kant’s Analytic-Synthetic Distinction: 
A Reply to a Classical Objection

Works by Others

The influence of Kantian philosophy as stretching to the present day is a well-established fact. One aspect of Kant’s work which has had a mixed impact is his insistence upon a duality of two types of cognitive content, that provided by concepts, and that residing in intuitions. It was rejected by post-Kantian German idealism and the Neo-Kantians. In the following, I would like to return to the original rejection of Kant’s concept/intuition dualism. This will shift the focus to consequences, in the foundations of mathematics. I shall look at the impact of the neo-Kantian discarding of the Kantian requirement for an intuitive content as essential in the representation of an object. The present paper has the limited scope of examining the impact upon Frege’s foundational work in the philosophy of mathematics, and proposing a Kantian interpretation of a way out of the problems it encounters.

Discarding Intuition: A Brief Historical Sketch

*Kant and the Post-Kantians*

That the discussion about intuition should have centred around perception is understandable insofar as the main interest is an account of our cognitive relation to a world of objects. But I would like to suggest that if one is to discuss Kant’s cognitive dualism concept/intuition, one should recall the key rôle that intuition plays in the Transcendental Deduction of the categories. According to interpreters ranging from Allison to Heidegger, the transcendental synthesis of the imagination plays a pivotal rôle in the deduction. The reason is that it is by providing a unity of the *a priori* forms of intuition that the categories apply to any manifold given in intuition. More
specifically, determinations of time are the product of this synthesis, and this is made possible by the schematisation of the categories. For our purposes, the main point however, is that it is only because there is some a priori form which characterises all intuition that a priori concepts such as the categories can be seen to have objective reality, i.e. in particular apply to manifolds given in outer sense.

Given the central importance of the Transcendental Deduction for the whole first Critique, the a priori forms of intuition would therefore seem to have a key rôle in Kant’s project. And the knowledge which these forms directly guarantee is, according to the Transcendental Aesthetic, mathematical, knowledge. The two consequences I draw from these observations are, first, that discarding the rôle of intuition would seem to amount to a rejection of a key tenet, rather than just a dispensable systematic feature of Kant’s project; and second, that discussing the importance of intuition must involve looking at mathematical knowledge rather than focussing exclusively upon perceptual knowledge.¹ These remarks will guide the investigation in this paper.

The starting point to an examination of the consequences of discarding the duality intuition/concept must be a historical survey of its sources. These lie in German Idealism. The search for an ultimate grounding of all knowledge which triggered Fichte’s early work involved an attempt to provide an unconditional first principle. Such unconditionality was not compatible with the endorsement of any dualism lying beyond the scope of the foundation in the first principle. For Fichte, the I as infinite activity produces the not-I, which forms the objectivity against which the finite I becomes conscious of itself. The I must, as a result, understand itself as finite to become aware of its infinite activity, and to grasp itself as infinite to understand its finitude. For Fichte, it is the imagination which enables these seemingly irreconcilable conceptions to be held together. Transcendental imagination, as for Kant, mediates between the understanding and sensibility. For Fichte, it grasps the infinite activity through its inhibition characterised by an intuitive manifold presented to the senses. In turn, this manifold provides that upon which the spontaneity of the understanding can exert itself. So, in effect, the duality between concept and intuition is transformed into a duality of perspectives in the I’s self-understanding, a duality which is mediated by the imagination. Kant’s unquestioning acceptance of a duality of cognitive contents is thus overcome within a system which absorbs it into a duality of perspectives upon the nature of the I, both unconscious and infinite, and conscious and finite. Such a conception of the imagination is adopted by Hegel who sees it as the source of the duality subject/object, and thus as identifiable with Reason itself. Within such a system, there is no place for a proper duality concept/intuition.

¹It must be mentioned that there is a default suspicion among British and American commentators in particular, towards Kant’s understanding of mathematics as knowledge involving intuition. This is largely due to the assumption that non-Euclidean geometry and formalism in mathematics must invalidate Kant’s views on the matter.
The Neo-Kantians

The neo-Kantians, both of the Marburg and the South-West schools, endorsed this post-Kantian rejection of a duality concept/intuition. Historically, this rejection was seen as essential to the rejection of psychologism which characterised both the neo-Kantian schools and Husserlian phenomenology, with its conception of “pure logic”.

It is interesting however to note the impact that the exclusion of intuition had upon neo-Kantian epistemology. In view of what we saw at the outset to be the rôle of the pure forms of intuition in accounting for the application of the categories, both neo-Kantian schools had to contend with the problem of providing a story for how the manifold given in sensation could be structured by pure concepts of the understanding. Since mathematics was viewed by Kant as the science of a priori intuition, we should expect the discarding of intuition to be reflected in a different understanding of the nature of mathematics.

Friedman (2000) describes the different approaches adopted by each school and how these reflect upon the understanding of mathematics. The Marburg school replaced the totally non-conceptual manifold of sensation with an already partly conceptualised manifold. This gets further conceptualised in the methodological progression of mathematical science. As a result, objective reality is defined as the limit of this progression. At any particular stage of the history of science, we only have a partial grasp of the object of knowledge. This amounts to a constructivist notion of what it is to be an object, and to a privileging of the knowledge provided by mathematical science. Such an epistemology of course requires that mathematics itself now be viewed as entirely conceptual, and thereby as belonging to logic. Such a view was made possible by the considerably broader conception of logic developed by the Marburg school, in comparison with Kant’s own notion of a science which is, ultimately that of syllogisms. In particular, for Natorp, logic includes the theory of relations. This enables Natorp to view the concept of number, which is based upon an asymmetric relation (that of the ordered series of numbers), as belonging to logic.

The Southwest school did not resort to this substitution of the manifold of sensation with a methodological progression of science. Rickert also insisted that the concept of number was not purely logical, although one must remember that the South-West school’s notion of logic was practically identical to Kant’s, i.e. much narrower than the Marburg school’s understanding of it. Nevertheless, the problem of accounting for the possibility of mathematics results from the gap left between the manifold of sensation and the logical forms of the understanding. More generally, the South-West school’s epistemology exhibits a duality between the realm of sensation and that of logic. The solution proposed by Emil Lask was to take the categories rather than the forms of judgement as primitive. Kant’s metaphysical deduction of the categories therefore becomes irrelevant: logic is rather now understood as derived through abstraction from the already categorised object. The categorised object is thus taken as defining the realm of value. But the problem of grounding value, i.e. of explaining the binding between the realms of being and of validity, a problem solved by Kant within his epistemological
framework by appealing to the transcendental subject, replaces the problem of the duality sensation/logic. The eventual outcome of this problem is that it will lead Heidegger to question the understanding of the subject and the timeless validity which lies at the heart of the philosophical tradition, a consequence which Friedman (2000, p.40-50) carefully unravels.

Russell’s Paradox

Having noted the broadening of Kant’s conception of logic by the Marburg neo-Kantians (Cohen, Natorp, Cassirer), we also note that the neo-Kantian inclusion of mathematics in logic is in tune with the logicist programme of grounding mathematics on the basis of logic alone. The need for a proper foundation for mathematics became an important preoccupation for mathematicians at the end of the 19th century. During that century, mathematics experienced many developments which spawned foundational problems: the construction of non-Euclidean geometries (Riemann, Lobachevsky), the theory of transfinite numbers (Cantor), … Hilbert was thus led to work on the foundations of geometry in the 1890s and later, in the 1920s, develop a formalist approach to grounding mathematics.

The initial work on grounding mathematics, however, saw it as the task of deriving mathematics from another theory, namely that of logic. The inspiration for this task lies in the development of the 19th century work by Bolzano on the grounding of the sciences upon “pure logic”. At the end of the 19th century, Frege published his work on the foundations of arithmetic. His ambitious attempt to derive arithmetic from logic however encountered a problem of inconsistency which Russell pointed out. Looking at this problem, I shall exhibit how it can be seen as a consequence of the ousting of intuition from epistemology.

This paradox is generally stated in terms of classes. There are classes which are not members of themselves. Thus, the class of all objects with the property of being a tree is not a tree. There are also classes which are members of themselves. Thus, the class of all objects which have the property of not being a tree is not a tree. Consider the class of all classes which are of the first type, i.e. not members of themselves. Is this class of the first or second type? If it is a class which is not a member of itself, then it is not one of the classes which are not members of themselves, i.e. it is a member of itself. But if we assume it is a class which is a member of itself, then it is one of its members, i.e. a class which is not a member of itself.

The paradox exhibits a problem for Frege’s axiomatic insofar as concepts have been exhibited for which one cannot define their extension. In an appendix to the *Grundgesetze*, Frege asks what are the implications of Russell's paradox: ‘is it always permissible to speak of the extension of a concept, of a class? And if not, how do we recognize the exceptional cases? Can we always infer from the extension of one concept's coinciding with that of a second, that every object which falls under the first concept also falls under the second?’ (Frege, 1964, p.127).
Standard Resolutions of the Paradox

If one cannot assign either of two mutually ‘complementary’ (i.e. mutually exclusive and such that either one of these properties has to be instantiated) properties to an object, the first response to the paradox which ensues must involve questioning the nature of the property in question. The property of not being a member of oneself is a property which involves a reference to the object itself. It is a kind of property that is predicated of the object qua class, namely that this class contains the object itself. This is the self-referentiality which is at the core of the paradox.

One option to avoid the paradox therefore consists in forbidding such self-referentiality. The most famous solution to the paradox has adopted this approach. This is Russell’s theory of types. Russell avoids such classes as those that are not members of themselves by arranging propositional functions (i.e. sentences of the theory) into a hierarchy. At the bottom of the hierarchy are sentences about individuals, and as we go up the hierarchy, we find sentences about classes of such individuals, then sentences about classes of classes of individuals, and so on. The theory of types then stipulates that it is only possible to construct classes which contain elements of one type (Russell, 1996). Therefore the question of whether a class contains itself is meaningless in such a theory. This first of all blocks the possibility of classes that are members of themselves. Second, the class of all classes that are not members of themselves is therefore the class of all classes and it is, because of its type, not a member of itself, which dispels the paradox. This solution works, but at the cost of a somewhat ad hoc subdivision of entities into different types. Moreover, it does not reflect the important fact that we can make sense of a class that is a member of itself. That is, if asked whether a class of certain objects is a member of itself, I am always able to answer the question, as long as I understand how membership of the class is characterised. Such an understanding and the concurrent ability to distinguish circumstances in which the property in question is instantiated or not, are not accounted for by Russell’s theory.

Quine (1980) proposes a way out of the paradoxical situation, which does not thus divide entities, and therefore provides a response to the latter criticism. Quine imposes restrictions upon the classes that are allowed to exist. These are namely those whose definition does not involve an infringement of the rule introduced in Russell’s type theory. That is while ‘not being a member of itself’ is meaningful, it leads to a defining condition for a class: ‘class of all classes that are not members of themselves’ which is not acceptable in the context of type theory, and therefore such a class does not exist in Quine’s system. The criticism of ad hoc-ness does not, however, disappear as a result of this Quinean variation on Russell’s theory of types.

The intuitionist approach in mathematics (van Stigt, 1990) does provide an account of why the property of including all classes that are not members of themselves, is to be rejected as definitive of a proper class. This is simply because no constructive process can be exhibited which yields such an entity. So, in effect, the intuitionist makes it a condition of objectivity that the object of the theory be constructible and Russell’s paradox is not a problem for him.
Such an approach is broadly Kantian in that it takes on board the Kantian idea that objectivity is a construct in which a concept is applied to a manifold. Although it does not refer to the Kantian notion of intuition as such, there is a sense in which the construction is a process which is in intuition insofar as it can be exhibited.\footnote{More would have to be said about the broader notion of intuition which is implied by the intuitionist move.} It does not, however, provide any further diagnosis of why it is that a property such as ‘is not a member of itself’ does not allow a class to be constructed out of it. The intuitionist solution locates the problem in the impossibility of using this property in a constructive process: this points to a problem in the property itself, but this problem is not thereby identified.

This is not however a shortcoming as such if the intuitionist replies that this is his way of dealing with the notion of objectivity, i.e. an equating of objectivity with that which is constructible in some sense. But such a notion of objectivity is distinct from, and of more limited applicability than Kant’s. Indeed, although in mathematics, one might want to accept the identification of the possibility of having an intuitive representation with the possibility of constructing this representation,\footnote{Further, we will see that this is more generally the case for classes of objects.} the intuitionist’s criterion refers to actual, not only possible constructions.\footnote{This leads of course to the intuitionist’s claim that what has not been shown to be true is neither true nor false.} Unlike Kant’s view, this does not address the fundamental question of what is an object, but only identifies objects constructed out of others, thus relying upon certain objects being given. Whether this is appropriate for mathematics is a question that cannot be addressed here. But if taken as a general characterisation of what can qualify as an object, the view that it has to be constructible out of more basic elements which are unproblematically viewed as having objective status is a view that is closer to phenomenalism (e.g. Ayer, 1978) than to Kant. On such a view, an object is the result of a construction out of the basic objects which are the sense-data we have access to in perceptual experience. This contrasts with Kant’s view that ‘the postulate for cognizing the actuality of things requires perception (...) – not immediate perception of the object itself the existence of which is to cognized, but still in connection with some actual perception in accordance with the analogies of experience, which exhibit all real connection in an experience in general’ (A225/B272). So, in summary, while objectivity for Kant lies in the possibility of an intuitive representation of an object that can be conceptualised in conformity with the principles of the pure understanding, the intuitionist approach relies only upon the actual conservation of the property of being an object through processes of construction.\footnote{As a result, an object which could (even in principle) never be directly perceived, such as a neutrino, is an object which clearly has a place through laws of empirical connection in Kant’s theory, whereas its existence for a phenomenalist is less clear. Note that the neutrino’s place in the objective world is part of a theory and arguably involves regulative principles.} With this more limited notion of objectivity, the intuitionist only provides a limited diagnosis of why the property of ‘not being a member of oneself’ leads to a paradox.
A Kantian Approach

The approaches considered above took as their starting point the view that, if a paradox arises from the impossibility of assigning one of two ‘complementary’ properties to an object, then the problem must lie exclusively in the property. But the solutions led to questioning the nature of the object (classes of all classes that are not members of themselves) itself insofar as this object is defined in terms of this very property. The other approach would therefore consist in directly questioning the nature of this object. This is a Kantian approach insofar as it examines the conditions under which there is an object of knowledge. The resolution which this approach leads to will necessarily be closely related to that proposed in the above realist approaches, for the reason noted that questioning the appropriateness of the property of ‘not being a member of oneself’ leads to questioning the nature of this object. But its motivation is fundamentally different.¹

I would like to suggest that it is possible in this way to formulate a solution to Russell’s paradox which draws directly upon Kant’s understanding of objectivity. In attempting to define an object as something characterised by its being a class and by a further property specifying what kind of class it is, it is the faculty of understanding which is at work in bringing a manifold of representations together. Intuitive representations of red objects can, for instance, be brought under a concept of class-hood. These intuitive representations are conceptualised intuitions (e.g. red objects with certain shapes and sizes, etc), and by bringing them under the concept of class-hood, one is viewing their property of forming a group.

The error of metaphysics, according to Kant, involves making knowledge claims for such cognitive syntheses in the absence of any intuition. This leads to making judgements, the truth-value of which cannot be assessed as they lie beyond all possible knowledge, i.e. the kind of knowledge which is available to a being endowed with a sensible intuition. Importantly, this is not to say that such syntheses of conceptual contents are meaningless. Kant makes it clear that I can talk of the self as having certain properties, even though I have no intuitions of the self.

When we consider the properties ‘being a member of oneself’ and ‘not being a member of oneself’, it is obvious that no intuition is available for such a property. This follows from the fact that our intuition is spatio-temporal, and that an intuition of a member of a class will have to be a spatial part of the intuition of the class, and therefore cannot also be that class itself. Note that this is the ‘intuition’ (in both senses of the word) which underlies Russell’s theory of types, although Russell would not acknowledge the spatial nature of such an insight.² As a result, to make the claim that I know objects which are defined as classes that are not members of themselves, is to make a claim which is beyond the bounds of possible knowledge. Even though the

¹Note that the transcendental realist has no tools at her disposal to ‘question the nature of objects’, and hence would seem to have to consider some more or less ad hoc ways of limiting the applicability of the self-referential property ‘not being a member of oneself’.
²There is no space here to discuss the sense in which I take even formal mathematics to involve spatio-temporal intuition. I refer to Parsons (1982), Hanna (2002), and Sutherland (2006).
understanding can grasp the synthesis of the concepts of class-hood and the
property of not being a member of itself, it can have no knowledge of objects
defined by a synthesis of these two concepts, because there is no intuition that
will enable the synthesis of these concepts. This does not, of course, prevent us
from having a grasp of the property ‘not being a member of oneself’ such that,
when an object is given, i.e. an intuition is already available conceptualised
(under another concept), one can evaluate the truth-value of the claim that it
has that property. Hence, one can claim that it is true of the class of all tea-cups
that it has such a property.

The problem of metaphysical thinking as Kant presents it, is that the faculty
of reason ignores these restrictions upon possible knowledge. It does so in
specific ways which are connected with transcendental illusions that lead the
faculty of reason to carry out hypostatizations of the contents of concepts when
no intuition is available. These transcendental illusions are themselves
unavoidable – e.g. the illusion that the ‘I’ of apperception refers to an entity
(Grier, 2001; Allison, 2004).

In the case in hand, the illusion which holds us in its grip is that if a
predicate is defined, it is possible to use it to define an object. I shall call it the
formalistic illusion. It is not necessarily a widespread illusion in the sense of
those identified by Kant, but it can be seen to have emerged in mathematical
logic as a result of the extension of logic from Kant’s narrow syllogistic
conception of it to include quantification in particular, as we shall see in more
detail below. In the case in hand, the faculty of reason considers the property of
‘not being a member of oneself’ as definitive of an object. As a result, it can
consider the class of all individual objects defined by this property, and this
turns out to be a problematic entity.

The Transcendental Realist Solution

If one wants to object that, in the light of the problems encountered with
this particular property, one would want to start with given objects and
segregate those which have the property, then a class of objects has to be given
at the outset, and these objects have to be classes. I can then form the class of
all these objects which, as classes, are not members of themselves. Such a
class, however, cannot be a member of itself because it is the result of a
construction which starts with objects (which are classes), and moves up to
classes of such classes in a way which excludes the possibility of the resulting
class actually being a member of itself. That is, it is the construction of a class
of such objects as distinct from the objects used in the construction which
excludes its having the property of containing itself. This way of getting rid of
the paradox constitutes the transcendental realist response to Russell’s
paradox, insofar as it assumes that objects are given at the outset.

Insofar as it relies upon constructions of classes of objects which are
distinct from the objects belonging to the class, the key may lie either in its
being a construction, or in its specifying classes as distinct from the objects
they comprise. The first option is, as we saw, that of the intuitionist
approach to the problem. It has the disadvantage of constituting a requirement upon all
“new” objects that is not grounded in an examination of objectivity (as Kant’s is), and leaves it open why the objects used as building blocks are exempt from such requirements.

The second option shifts the focus from the constructive nature of the constitution of new objects to the different types of objects thus generated. If a class is of a different type from one of its elements, the problem is averted. Russell’s (and Quine’s) solution is to introduce a hierarchy of types, so that what is described as a construction above amounts to a move to a higher type. The problem with this is that it only works as a solution if one starts off with objects of the same type. That is, when considering the objects as classes of which we isolated those which are not members of themselves, if Russell’s theory of types is to achieve its result of dispelling the paradox, it requires that the classes that are referred to are themselves of the same type. So, although this approach is appealing as a natural transcendental realist response to the problem in hand, it is ad hoc in that the solution it proposes relies upon the objects which are given at the outset having the kind of property specified in the solution.

The Transcendental Idealist Solution

To understand in Kantian terms how a paradox arises, one must first note that it draws directly upon logical tools which were not available to Kant. Although Kant may have had the notion of universality defined as a function of unity of judgements, he did not have access to the quantifier $\exists\exists$ in the variety of its uses. In particular, quantification allows for the definition of classes through predication, i.e. as $C = \{x \mid f(x)\}$ or $(\forall x)(x \in C) \leftrightarrow (f(x))$. This Fregean development of logic\(^1\), which follows upon Cantor’s work in the theory of sets, leads to the construction of entities of which transcendental philosophy asks under what conditions they are objects.

I claim that a formalistic illusion arises as a result of reason’s demand for completeness of the conditions under which anything is given, and that this illusion amounts to treating all such entities as objects. Let us recall that Kant refers to the function of reason in terms of its ‘transcendental concept’ which is “the totality of conditions to a given conditioned thing” (A322/B379). If one considers an object $x$ to which the predicate $f$ applies, one can form the idea of the class $F = \{x \mid f(x)\}$ of all objects with this property, i.e. the idea of the extension of $f$. And this class can be viewed as a condition of the conditioned object $x$.

There is however a difference between this formalistic illusion and Kant’s hypostatization of transcendental ideas. Unlike Kant’s ideas of God, the soul and the world, the extension of many concepts can form a perfectly unproblematic object of knowledge. Thus, the class of all trees is such an object. Why? Because rather than viewing it as a totality of conditions, I would

\(^1\)It is one of many developments arising from the formalisation of logic, but it has implications that go further than purely formal innovations such as predicates with several places.
know in principle how to go about identifying it, i.e. through a process by which all the manifold in intuition is brought under the concept ‘the totality of all trees’. The availability of such a process is ensured by construction, from the objects themselves to the class of all such objects.1 In the case of self-referential predicates, such a constructive process is not available. In such cases, the formalistic illusion leads to defining something which is not an object. In line with what we noted above, although we find a resolution of the paradox related to the intuitionist’s, it is not construction per se which is here definitive of what counts as an object, but the possibility (in principle, that is, in line with the principles of the pure understanding) of some intuitive representation which is conceptually determined as an object (see A156/B195). In the case of classes of objects, the possibility of a construction of the class coincides with the possibility in principle of an intuitive representation of the class.2

This resolution of the paradox has the advantage of accounting for the fact that one is able to understand the property N: ‘not a member of oneself’, even though one is not able to represent such a property in intuition. The distinction concept/intuition enables us to see that this understanding is conceptual: as long as an object x is already given, I can know the truth-value of Nx. For such an object, I can have an intuition of it under another conceptualisation, and must then only examine whether this intuition can be brought under concept N. But there is no intuitive representation of an object defined solely by property N, which gives rise to the formalistic illusion.

This also shows why a transcendally idealistic perspective is required to resolve the paradox satisfactorily. That is because the paradox results from ignoring the conditions required for something to be an object, i.e. the possibility of an intuitive representation of a manifold determined by a concept. With self-referential properties, the grasp we have of them relies upon an object already being given. By ignoring this fact, the construction of the extension of such a property does not yield an object.

References


1The issue is not the possibility of perceiving such an object, but of knowing how it can be identified, and this is possible through a constructive process.
2The issue of our actual ability to intuit the class in question is a contingent matter, and as Kant notes ‘the crudeness [of our senses] (…) does not affect the form of possible experience in general’ (A226/B273)
In respect of the so-called doctrinal parts of his philosophy, Kant proposed strict grounds and rules for thinking. Otherwise, however, he was puzzled by the all-important question: *What, in the absence of strict rules, counts as intelligible thinking?*

By ‘intelligible thinking’ I refer to the reflective working of the human mind, which acknowledges no strict rules, but yet is able to orient itself, set ends for itself, and aim for objectivity. ‘Intelligible thinking’, hence, refers to a non-doctrinal but yet apt way of thinking.

I will first study the reflective core of intelligible thinking: the so-called architectonic unity of thinking (Section 1). Then I will consider the grounds and the nature of intelligible thinking, paying a special attention to the close relation of intelligible thinking to Kant’s aesthetics (Sections 2 to 4). Finally, I conclude my study (Section 5).

**Reflective Thinking: The Architectonic Unity of Reason**

The starting-point of any kind of intelligible thinking is the given experience about which reason aims to form a coherent understanding. The attempt to form an understanding of anything can, however, never even get started without some kind of systematic apprehension of the given situation. In this respect, Kant speaks of the architectonic tendency of our thinking – the so-called *architectonic unity of reason* (CPR, A474/B502; A832/B860 ff.) – which, in the absence of determinate guidelines, yet gives some kind of non-determinate coherence or unity to experience.¹

¹The idea of the architectonic unity of reason appears in Kant’s *Critique of Pure Reason* (1781/1787), where it is introduced as a mode of reflective thinking for the purposes of a philosophical system.
Kant sets two conditions for the architectonic system of thinking. First, the different parts of system cannot contradict each other. As Kant notes, reason ‘…regards all our knowledge as belonging to a possible system, and therefore allows only such principles as do not at any rate make it impossible for any knowledge that we may attain to combine into a system with other knowledge.’ (CPR, A474/B502). No system of thought may include elements that are in an outright contradiction with the other elements of the system. The second condition for the architectonic unity of thinking is, according to Kant, that there is some idea (or ideas) that orient our search for a system (even though these ideas cannot bring forth any specific system). Kant, accordingly, notes that system is …the unity of the manifold modes of knowledge under one idea. This idea is the concept provided by reason – of the form of a whole – in so far as the concept determines a priori not only the scope of its manifold content, but also the positions which the parts occupy relatively to one another (CPR, A832/B860).

The architectonic unity of thinking, hence, requires a (reasonably) non-contradictory account of the various elements of the given situation (the 1st condition), as well as some ideas or principles, which orient thinking in its effort to comprehend the situation (the 2nd condition).

Architectonic thinking is essentially reflective: the starting-point for the use of reason is the given manifold of knowledge out of which reason aims to form a systematic whole. Being reflective, as Susan Neiman notes, reason also essentially requires criticism and introspection (Neiman, 1994, 4-5). Since there is no established formula for the use of reason in its reflective task, it is at the heart of it to critically scrutinize its own conduct. This is essential to Kantian intelligible thinking.

Notably, Kant proposes various conditions for intelligible thinking: 1) the collective conditions, 2) the transcendental conditions, and 3) the sensible conditions (feelings).

The Collective Conditions

Intelligible thinking has socially embedded, collective conditions, which, as will be seen, prompt not only the ideas of societal and communal conditions of thinking - including two issues: 1) the sociability of a human being, and 2) the public use of reason - but also reveal the relation of intelligible thinking to Kant’s aesthetics, and show the historical nature of intelligible thinking.

For any kind of thinking to emerge, one, first, has to socialize oneself. Important as it is for man to think for himself, and not to be led by a common understanding - see also below, gemeiner Menschenverstand - his nature drives

---

1 We might also talk here of principles, as well as certain conditions, which orient our use of reason (see more below, Sections 2 to 4).

2 This same idea is also expressed in CPrR, 5:10.
him also to what Kant calls sociability (Geselligkeit).\(^1\) Kant defines sociability as the suitability and the tendency toward society (CPJ, 5: 296-297), and regards it as ‘…necessary for human beings as creatures destined for society, and thus as a property belonging to humanity…’ (CPJ, 5: 297). No single person can, according to Kant, realise her capacity for thinking - or, for that matter, humanity - alone, but always needs a social playground for her thoughts.

Closely related to the idea of socialization is Kant’s contention of culture as a background for the human ability to think. Human being’s capacity and predisposition for thinking cannot, according to Kant, be fulfilled without a proper cultural background. Only in the world of culture can one come to know and realize her (proper) ends (see more below).

The relevance of the cultural sociability for human thinking is clearly present in Kant’s account of aesthetic judgments. Kant maintains, as Paul Guyer notes, that a) our pleasure in the beautiful cannot exist in the absence of society, b) the existence of society is necessary to help us form our taste, and c) making universally valid judgments of taste is of interest to us only in society (Guyer, 2005, 380, n. 16). Though Kant’s consideration of aesthetic judgments (especially in the Critique of the Power of Judgment) is focused on their pure, a priori features, he never shuns the idea that without society no aesthetic judgments would occur.

For the present purposes, it is important to note that aesthetic judgments are Kant’s example of reflective judgments. If sociability and the existence of society are important for reflective aesthetic judgments, we may think that Kant sees them important also for other kinds of reflective judgments. Having recognized the reflective nature of intelligible thinking (see above), we may suggest that sociability and the existence of society are important also for intelligible thinking.

The second collective condition for intelligible thinking Kant calls the public use of reason. To socialize one-self is not enough for the emergence of intelligible thinking. One also has to think publicly. By the public use of one’s reason Kant understands ‘…that use which someone makes of it as a scholar before the entire public of the world of readers.’ (WE, 8:37). Kant’s idea of the public use of reason, as Paul Nnodim notes, entails an inner ability and courage to take up the task of thinking for oneself in public, rather than allowing others do it on one’s behalf (Nnodim, 2004, 149-150).\(^2\)

Besides this emphasis on a personal courage, the idea of the public use of reason prompts the idea of communication as a precondition of apt thinking. This has a significant relevance to intelligible thinking. Kant holds that, without free communication between citizens, the old and the new prejudices of society retain, and render citizens passive and unable to use their reason. It appears, indeed, that Kant proposes communication as the very metaphor for apt thinking or use of reason. This is, at least, if we turn to the Critique of Pure Reason where Kant, in his motto to the second edition of CPR, in the words of

\(^1\)On Kant’s concept of unsocial sociability, see IUH, 44-45.

\(^2\)On Kant’s idea of the private use of reason, see WE, 8:37.
Francis Bacon asks people to ‘…discuss together…’ (CPR, B ii; translation Onora O’Neill, 1989, 6), and where he presents an image of reason as a debate between free citizens. Kant writes:

For reason has no dictatorial authority; its verdict is always simply the agreement of free citizens, of whom each one must be permitted to express, without let or hindrance, his objections or even his veto (CPR, A738-9/B766-7).

I suggest that Kant’s metaphor of the debate between the free citizens (CPR) and the idea of the public use of reason (WE) yield the same idea: apt or intelligible thinking essentially depends on - though is not reduced to - the soundness and communicability of the different points of view proposed by actual reasoning human beings.

To support the thesis that the question here is precisely of an actual - that is, societal and empirical – communication, we may turn to Kant’s note that if thinkers ‘…exhibit talent, if they initiate new and profound enquiries, in a word, if they show reason, reason always stands to gain.’ (CPR, A746/B774). Reason indeed stands, according to Kant, in need of a dialectical debate instituted with unqualified public approval (CPR, A747/B775). This demand for the collectivity of thinking is also present in Kant’s note that ‘…the same external constraint which deprives people of the freedom to communicate their thoughts in public also removes their freedom of thought…’ (WOT, 247). The individual ability to think is dependent on the possibility to share one’s thoughts with others.

Now, this account of thinking is historical – that is, bound up with the actual debates between (preferably enlightened) citizens. The question of the historical nature of Kant’s philosophy is often, however, and quite understandably, either denied or paid no attention. It is, after all, a fact that much of what Kant says about both theoretical and practical uses of reason allows no historical interpretation. In respect of the historicity of Kant’s account of human thinking/reason, it is, however, worth considering Kant’s own account of history, which, at least, acknowledges the idea of the historical development of reason.

In this respect, Mary J. Gregor notes that Kant views history as the account of what nature does to prepare the human race for its final end. Taking the human race collectively, nature impels the human being toward a state in which she is ready to realize her capacity and predisposition for freedom and, ultimately, for moral freedom. Gregor also notes that, for Kant, the state of civil society provides the framework within which the ‘seed of good’ inherent in the human being can develop (Gregor, 1974, xxii.) Lewis White Beck in turn maintains that for Kant history is the recounting of the movement of a human being from the state of being a part of the mechanism of nature to the state of being the creator of the world of culture, where she can eventually come to know and perform her duties and realize her moral ends (Beck, 1963, xviii). Onora O’Neill, for her part, claims that Kant’s developmental account of reason fits well with an evolutionary view of cognition (O’Neill, 1989, 43), and
that, in Kant’s view, actual human reasoning occurs in and must to some extent be premised on human institutions (O’Neill, 1989, 17).

Kant then allows the idea of the historical development and realisation of human reason, and also recognises the role of collective institutions as (necessarily historical) preconditions for human thinking. O’Neill even notes that Kant’s Critique of Pure Reason can itself be seen as an episode in the history of reason. The historical task of finding the right plan for using the ‘materials’ that evolutionary pressures have produced is still incomplete, and so Kant seeks fellow workers and further effort (O’Neill, 1989, 22.) Beck, on his behalf, notes that for Kant culture is not only the fruit of reason. It is also a condition for the perfect growth and full employment of reason itself (Beck, 1963, xxiii.)

On this basis, I claim that, according to Kant, the full development of intelligible thinking has as its preconditions the fulfillment of the cultural sociability of a human being, and the freedom for public communication (the public use of reason).

The Transcendental Conditions

Certain transcendental conditions,¹ Kant maintains, also orient intelligible thinking. The so-called principles of purposiveness (Zweckmässigkeit) and sensus communis (sensus communis, Gemeinsinn) must, at least, be included among these conditions. Both of these principles also show a remarkable affinity with Kant’s account of aesthetic judgments.

The principle of purposiveness is, according to Kant, fundamental to any kind of reflective thinking. The basic function of this principle is somewhat the same as that of the architectonic unity of thinking: the bringing of unity to where no unity is given. The principle of purposiveness gives unity to experience by suggesting a common point of reference (a concept), under which the different elements of experience can be subjected. By applying the idea (or a concept) of a purpose to nature and objects, the principle of purposiveness allows us to orient ourselves in otherwise quite multiple experience.

There are, according to Kant, two kinds of purposive judgments: objective and subjective. Since the objectively purposive judgments have more closed structure, which makes their application to the idea of intelligible thinking more difficult, I will concentrate on the subjectively purposive judgments.

Kant, notably, identifies subjectively purposive judgments with aesthetic judgments, which according to him, set objects ‘…the purposiveness…without representation of an end [ ]’ (CPJ, 5:236), providing us with concepts in the face of the excessive multiplicity in nature, in order to be able to be oriented in it (CPJ, 5:193).

The purposiveness of aesthetic judgments is, according to Kant, free: no doctrinal guidelines, determinate concepts, or pre-given purposes orient

¹By ‘transcendental conditions’ I refer to a priori – contra historical or empirical – conditions of human thinking.
aesthetic judgments (see above: ‘…the purposiveness…without representation of an end [ ]’). This, however, does not mean that aesthetic judgments would lack all orientation. Though no determinate concepts – which, according to Kant, give direction both to theoretical and practical (determinate) judgments of reason – orient aesthetic judgments (CPJ, 5:217), still some other kinds of concepts do. Samuel Fleischacker notes that it is not that in a Kantian aesthetic judgment the cognitive powers are not limited by any concept, but that no determinate concept limits the free play of the faculties. To think without a determinate concept might, according to Fleischacker, instead be understood as using concepts without defining precisely which concept one is using: to allow a range of concepts to play with the contents of one’s imagination instead of fixing one of them determinately to that content (Fleischacker, 1999, 23-24.)

Though no determinate concepts determine aesthetic judgments, they are nevertheless ‘determined’ by ‘non-determined’ concepts. This gives aesthetic judgments orientation, yet allowing their free functioning. This interplay between freedom and determination is essential also for intelligible thinking. While there are no doctrinal guidelines for intelligible thinking, there yet are regulative guidelines for it.

The principle of sensus communis, according to Kant, is also important for intelligible thinking. Kant describes sensus communis as

...a communal sense, i.e., a faculty for judging that in its reflection takes account (a priori) of everyone else’s way of representing in thought, in order as it were to hold its judgment up to human reason as a whole and thereby avoid the illusion which, from subjective private conditions that could easily be held to be objective, would have a detrimental influence on the judgment (CPJ, 5:293-4).2

Sensus communis, then, is a principle, which, by constituting the possibility of a general - or, in a sense, ‘neutral’ – viewpoint, allows human beings to overcome their (wrong kind of) individuality.

In this sense, by generating the possibility of the general, sensus communis can be seen to be presupposed in every judgment aiming at communicability. Kant notes that

---

1Kant’s so-called aesthetic ideas are important in this respect. Kant notes that aesthetic idea is ‘…that representation of the imagination that occasions much thinking though without it being possible for any determinate thought, i.e., concept, to be adequate to it, which, consequently, no language fully attains or can make intelligible.’ (CPJ, 5:314). Aesthetic idea might then be characterized, as Kirk Pillow suggests, as a concept aesthetically expanded in an unlimited way by bringing its implications and its kinships with other concepts to bear on its rich meaning (Pillow, 2001, 200), or, as Steven Ravett Brown suggests, to a certain extent as a kind of overflow of content (Brown, 2004, 489).

2Kant makes a distinction between the concepts of sensus communis/Gemeinsinn and gemeiner Menschenverstand. The concept of gemeiner Menschenverstand (common human understanding) Kant understands as merely healthy - not yet cultivated – understanding, that is, the least that can be expected of anyone laying claim to the name of a human being (CPJ, 5:293).
...a common sense [Gemeinsinn]...must be able to be assumed...as the necessary condition of the universal communicability of our cognition, which is assumed in every logic and every principle of cognitions that is not skeptical (CPJ, 5:239).

Hence, as Rudolf A. Makkreel notes, Kantian sensus communis uses reflective judgments to abstract from the private empirical aspects of our subjective representations in order to generate what might be called a communal or inter-subjective perspective (Makkreel, 1994, 158). The principle of sensus communis, hence, has somewhat the same role as the public use of reason: by introducing a communal and communicative point of view, the principle of sensus communis helps apt thinking. In this role, it is crucial also for intelligible thinking.

In respect of the principle of sensus communis, too, Kant refers to aesthetic judgments. As noted, Kant makes a strong case for the fact that aesthetic judgments are both free and in need of orientation. While the freedom (from strict rules) is a necessary precondition of aesthetic judgments, so is the communicability of these judgments. It is the task of the principle of sensus communis to generate this communicability - that is, to make aesthetic judgments collectively meaningful, or to make possible the demand for the acceptance of others of our aesthetic judgments.

Though aesthetic judgments, as concerning aesthetic feelings (see below), are, according to Kant, subjective, they also have the peculiar feature of demanding objectivity: by making aesthetic judgments, we demand that our fellow beings accept them. This kind of search for objectivity is characteristic also for intelligible thinking: in its aim at common acceptability, intelligible thinking can’t rely on any doctrinal guidelines. This means that another kind of (non-doctrinal) back up is needed if intelligible thinking is to gain objectivity. This will be discussed below (Section 4).

In conclusion, then, both the principle of purposiveness and the principle of sensus communis are essential for intelligible thinking in its task to orient itself in multiple experience.

The Sensible Conditions

Besides the celebrated a priori nature of a human being, man also has, even according to Kant, a sensible existence. This, I suggest, has a role to play in Kant’s account of intelligible thinking. Kant’s considerations of the feeling of pleasure (Lust) and the feeling of life (Lebensgefühl) in an aesthetic experience are interesting in this respect. By advancing aesthetic experience, these feelings help aesthetic judgments - and, so I argue, also intelligible thinking - in their search for objectivity.

The aesthetic feeling of pleasure is, according to Kant, intimately linked with the free play of the faculties of cognition in an aesthetic judgment. Kant holds that when the faculties of imagination (Einbildungs Kraft) and understanding (Verstand) acquire a state of mutual harmony, there appears the aesthetic feeling of pleasure. This is a specific kind of pleasure: it concerns the
unification of two contrary faculties of mind. While imagination gives strive and impetus to an aesthetic judgment, and recognizes no preset boundaries, understanding, by limiting the work of imagination, offers an orientation to the aesthetic judgment. This work of understanding was above referred as the ‘non-determined determination’ of aesthetic judgments. When understanding non-determinately and harmoniously orients imagination, there emerges the aesthetic feeling of pleasure. As noted above, this kind of interplay between the determinate and the free aspects of thinking is essential also for intelligible thinking.

The feeling of pleasure again, Kant holds, enhances the feeling of life. Together they make aesthetic experience an ongoing business. Aesthetic feelings give aesthetic judgment endurance, and make possible its ongoing pursuit for objectivity (or its effort to make collectively meaningful and acceptable judgments).

Intelligible thinking, I argue, acquires objectivity in a similar manner. To begin with, both aesthetic judgments and intelligible thinking start their pursuit for objectivity in a like manner. Supported by the above-mentioned conditions of reflective thinking (the architectonic unity of reason, as well as the collective and the transcendental conditions of thinking) they aim at common acceptability. Without the feeling of pleasure and the feeling of life, the specific objectivity of both aesthetic and intelligible judgments would, however, remain in a quite preliminary stage. Whatever kind of objectivity there is to be achieved, it can’t be earned if we are not able to enjoy our experiences/thoughts related to our judgments (through the feelings associated with them). In order to keep up the search for objectivity, we need pleasurable feelings. Anyone ever committed to figuring out something in the absence of explicit guidelines for one’s search, should not find it surprising that a certain element of pleasure is needed in order to have the strength to keep up the search.

I agree with Howard Caygill that, with Kant, the question of feeling relates closely to the problem of the relationship of philosophical speculation to areas of human life and activity which do not obey limited or formal notions of rationality (Caygill, 1995, 199). I only add that the above-mentioned questions concerning the sociability of a human being, the public use of reason, and the transcendental principles of purposiveness and sensus communis also relate to this matter. They all ground intelligible thinking.

**Conclusion**

In this paper, I have argued that Kantian intelligibility is a broad and non-doctrinal brainwork, which shows an intriguing affinity with Kant’s aesthetics. Intelligible thinking is based on the idea that human thinking/reason is, in a certain sense, both free and preconditioned: while there is no doctrine for thinking, there yet are grounds for the orientation in thinking. Some of these grounds have been described above.

Throughout the article, I have emphasized the relevance of Kant’s aesthetics for intelligible thinking. It is reasonable, I claimed, to widen Kant’s
account of one species of reflective judgments - that is, aesthetic judgments – to count also for other kinds of reflective judgments, as, for example, those of intelligible thinking.

I also showed how closely the collective, transcendental and sensible conditions of intelligible thinking relate to Kant’s account of aesthetics. It remains to be shown that also Kant’s account of the architectonic unity of thinking bears a remarkable affinity with his aesthetics. The two above-mentioned demands of the architectonic unity of thinking, namely, are met with Kant’s aesthetic judgments. By ‘playing’ with various viewpoints or conceptions of objects, Kantian aesthetic judgments meet the first demand of the architectonic unity of thinking - that is, the demand to bring the various elements of the system under some non-determinate principles/ideas, which orient our thinking – aesthetic judgments meet by subsuming themselves to the principle of subjective purposiveness, or, as Kant says, to a ‘…lawfulness without law…’ (CPJ, 5:241).

Bibliography

Works by Kant

All the English translations of the German quotations are from the editions cited below. The references (except IUH, and WOT) follow the pagination of Kant’s gesammelte Schriften, herausgegeben von der Königlich Preußischen Akademie der Wissenschaften, Berlin: Walter de Gruyter (the so-called ‘Akademie’ edition).

CPR

IUH

WE

WOT
‘What is Orientation in Thinking?’ (Was heisst: Sich im Denken orientieren?) (1786), translated by H.B. Nisbet, in Kant: Political Writings, Cambridge Texts in the History of Political Thought,


Other Works


Throwing Out the Scholastic Metaphysics with the Bathwater?

T. Ann Scholl, United Arab Emirates University, United Arab Emirates

I. Well understood is Descartes’s split, both epistemically and metaphorically, from his Scholastic predecessors. Descartes certainly rejects Scholastic hylomorphism, the belief that a human is one substance in which the mind provides form to the body. (CSM I.213: AT VIII A.29) Descartes also rejects aspects of Scholastico–Aristotelian psychophysiology, particularly the metaphysical notions of phenomenal resemblance, and the transference of essences via the sensory systems, resulting in ideas of the corporeal objects sensed. (CSM I.165: AT VI.113) Some Cartesian scholars suggest that Descartes does away with the notion of representational sensory ideas altogether. In some places, Descartes perhaps waxes Platonic, suggesting that sensory ideas are innate. (CSM I.304: AT VIII B.359) Herein, I propose that Descartes retains some Scholastic notions of representation in sensory and imaginary ideas, despite his rejections of Scholastic hylomorphism and Aristotelian precepts and his possibly embracing the notion of innate sensory ideas.

II. To understand Descartes’s notion of representationality in sensory ideas, one must first examine how Cartesian psychophysiology of sensory perception differs from that of his Scholastic predecessors. While the Scholastics vary in the exact details of how sensation works, the basic presumption is that something, an essence in some form or other, is transmitted from the object of

1Cottingham, Studoff and Murdoch, editors. The Philosophical Writings of Descartes, Vol. I. p. 187. (Cambridge: 1984). Hereafter, references to Descartes’s works will be from this translation, referred to as ‘CSM’ followed by volume numbers and page numbers and followed by standard reference ‘AT’ to the Adam-Tannery volumes.
sensation to the mind via the external sense organs. In the mind, this essence becomes an intentional essence forming the core content of a sensory idea. This intentional essence is the basic core of a semblance which represents, as a picture—perhaps even an abstract one—might.

According to most Scholastic philosophers, a cognizer becomes aware of the external object via this semblance in the mind. Descartes, being more metaphysically thrifty than his predecessors, rejects the notion that intentional essences exist and that intentional essences float through mediums and arrive at the mind via the sensory organs. (Dioptrics, CSM I.164: AT VI.112) In Descartes’s metaphysical world, the physiology of sensation operates mechanistically. The nerves being stimulated by contact with external corporeal substances, then send animal spirits in various patterns to the pineal gland, which is in turn shaped and moved to stimulate the mind to form a sensory idea that is apprehended by the attending mind. The sensory organs and their actions are governed by purely mechanistic, causal rules. No need to introduce, Descartes believes, such unnecessary metaphysical entities intentional essences and quasi–phenomenal semblances.

Setting aside the notorious difficulty of how this mechanistic operation might produce a phenomenal idea, the difficulty becomes a question of how the phenomenal content of a sensory idea might make an attending mind aware of an external corporeal substance. Descartes clearly indicates in Dioptrics that sensory ideas do not resemble the external corporeal substances they present to a cognizer, saying

We must take care not to assume—as our philosophers commonly do—that in order to have sensory perceptions the soul must contemplate certain images transmitted by objects to the brain; or at any rate we must conceive the nature of the images in an entirely different manner from that of the philosopher...And if, in order to depart as little as possible from accepted views, we prefer to maintain that the objects which we perceive by our senses really send images of themselves to the inside of our brain, we must at least observe that in no case does an image have to resemble the object it represents in all respects, for otherwise there would be no distinction between the object and its image. It is enough that the image resembles its object in a few respects. ...You can see this in the case of engravings: consisting simply of a little ink placed here and there on a piece of paper, they represent to us forests, towns, people, and even battles and storms... (AT VI.113–4: CSM I.165–6)

Here Descartes asserts that sensory ideas are not, nor do they contain, semblances in the Scholastic sense: sensory ideas do not literally resemble the external corporeal objects. Descartes’s use of the Scholastic engraving analogy leaves open the possibility that Descartes accepts, perhaps in a fashion of signification, that sensory ideas represent the essence of a corporeal substance to the attending mind. This representation, however, must maintain differences from the object represented in an idea and is thus in many respects not a semblance at all. Most importantly, Descartes implies than an essence, a nature
Throwing Out the Scholastic Metaphysics with the Bathwater?

itself, is not contained within the phenomenal content of a sensory idea—instead, the essence is merely represented in a way that causes the attending mind to become aware of the represented object. This representation, in turn, makes the mind aware of the essence of the sensed external corporeal substance.

While this is a rejection of Scholastic notions of representations, it remains only a partial rejection of Scholastic notions. Descartes still allows for the possibility, if not necessity, of a phenomenal representation of some sort. What Descartes rejects in *Dioptrics* are Scholastic claims that corporeal essences are transmitted to the mind and that these essences form the core content of a phenomenal idea. At least in *Dioptrics*, he does not reject the basic Scholastic concept of representation.

III.

The question of essence complicates matters for Cartesian sensory ideas, especially in connection with the relation of the phenomenal representation of an idea to the essence of the object represented. The previously quoted passage from *Dioptrics* suggests that Descartes rejects the Scholastic notion that sensory ideas contain the essence of the sensed corporeal object, for Descartes clearly notes his wish to distinguish the phenomenal images from the objects they represent.

Today, we scoff a wee bit at Scholastic essences floating through the air and up our nose and onwards to the brain, as well as those mysterious hylomorphic powers. These essences, however, solve quite a bit for a representationalist. A purely phenomenal idea containing the representation of a corporeal substance seems self-contradictory, for instance. The most notorious problem is that of the physiological to mental (and *vice versa*) causation. By retaining representationality and rejecting the Scholastic operations and explanations of sensation in favor of a less complex metaphysical world, Descartes seemingly leaves himself with more metaphysical problems than not.

Of importance, however, is Descartes’s claim that the sensory ideas resemble the sensed object only in regard to shape. Ideas of shape, according to Descartes, are ideas of “true and immutable natures,” which are the subject of pure geometry, understood by the intellect alone or the intellect aided by the imagination. (CSM III.226: AT III.691–2) In *Meditation V*, Descartes asserts, while contemplating the possible existence of external corporeal substance that

…I find within me countless ideas of things which even though they may not exist anywhere outside me still cannot be called nothing; for although in a sense they can be thought of at will, they cannot be called nothing; for although in a sense they can be thought of at will, they are not my invention, but have their own true and immutable natures. When, for example, I imagine a triangle, even if perhaps no such figure exists, or has ever existed, anywhere outside my thought, there is still a determinate nature, or essence, or form of the triangle.
which is immutable and eternal, and not invented by me or dependent on my mind. (CSM II.44–5: AT VII.64)

This might suggest that Descartes accepts a Platonic viewpoint, that the mind apprehends a true and immutable nature, found innately within the mind, rather than gained from outside the mind. The above passage may only refer to geometrical shapes: those ideas of corporeal objects that are the subject of “pure–mathematics.” If so, the concept of representation in Descartes’s views of sensory perception would differ greatly from that of the Scholastics.

In Comments on a Certain Broadsheet, Descartes further remarks that

…not because these things transmit the ideas to our mind through the sense organs, but because they transmit something which, at exactly that moment, gives the mind occasion to form these ideas by means of the faculty innate to it. Nothing reaches our mind from the eternal objects through the sense organs except certain corporeal motions…The ideas of pain, colors, sounds and the like must be all the more innate if, on the occasion of certain corporeal motions, our mind is to be capable of representing them to itself, for there is no similarity between these ideas and the corporeal motions. (CSM I.304: AT VIII.B.359)

Here is where Descartes seemingly throws out the Scholastic metaphysics with the bathwater, suggesting that nothing—i.e., no essence of any kind—transmits itself to the mind via the physiology of sensory perception, thus following those (de Clave, Basso, etc.) expelled from Paris for anti–Aristotelian viewpoints. Yet it is not clear that Descartes rejects, in full, the Aristotelian viewpoint regarding sensory ideas. His notion of representation appears, in Dioptrics at least, to retain somewhat of the Scholastic mindset beyond mere retention of the Scholastic lexicon.

While Descartes rids himself of cumbersome Scholastic metaphysics regarding psychophysiological processes of sensation, he still suggests that the mind can apprehend the essence, perhaps even suggests that the ideas of corporeal substances contain some essence—one that is perhaps not transmitted from sensory physiology to sensory phenomenal capacities. His assertions in Comments on a Certain Broadsheet may simply indicate that the mind is innately able to respond to certain corporeal motions in the brain (more specifically, the motions of the animal spirits as they strike and rebound from the pineal gland), but that such ideas require the a physiological stimulation and active creative causation. Descartes clearly indicates that sensory ideas are caused by physiological motions. This causation, however, is not causation in the sense that the phenomenal content itself is transferred to the mind. The mind, given that it has the phenomenal power of sensory perception is innately

---

set up to “receive” the motions of the animal spirits. ¹ This is not necessarily an indication, by itself, that the idea is innate, merely that the mind is, by design, set to be affected by the sensory ideas. The notion of creative causation, might help explain. The sensory system, while transferring no phenomenal content to the mind, nonetheless determines, i.e., causes, the content of a resulting sensory idea. So, when encountering an object which, all things being equal, produces a sensation of red, the sensorial physiology does not transmit the idea of red, but instead creates or causes an idea of red in the mind. One need not claim a Platonic, innate idea in this case.

IV.

So, even now, it is not clear that Descartes rejects the Scholastic concept of representationality. Consider Descartes claims concerning the ideas of “pure mathematics:

…when I was completely preoccupied with the objects of the senses, I always held that the most certain truths of all were the kind which I recognized clearly in connection with shapes, or numbers or other items relating to arithmetic or geometry, or in general to pure and abstract mathematics. (AT VII.65: CSM II.45)

This refers to those “clear and distinct” ideas of the shapes of pure mathematics which are understood by the mind as well as perceived by the senses or pictured in the imagination. The relevant difficulties are a) whether Descartes accepts that ideas that are about the subject of pure mathematics contain an essence and b) whether sensory ideas also contain these essences.

In regard to the first difficulty, it is not clear that Descartes merely rejects the existence of Scholastic intentional forms and not the concept of representation of essences. In Principles, Descartes clearly lays out his epistemology based on the Scholastico–Aristotelian concept of principle attributes. A substance, he clearly states in Principles and elsewhere, is known through any one of its attributes. ² Elsewhere, I have argued that for Descartes, modes of a substance are functions or operations of a principle attribute. Accordingly, a mode of a substance is merely the principle attribute itself, in a particularized mode of function or operation. Cognitive awareness of an attribute or mode of a substance is cognitive awareness of the principle attribute—the core essence—of a substance. Hence, to become aware of a mode of a substance it to become aware of the substance’s principle attribute, or its essence. Therefore, for Descartes, becoming aware of a corporeal substance is to be aware of its principle attribute or essence. The idea of a corporeal substance must then, in some way, present a principle attribute to an attending mind.

¹In many places, notably Meditations III and VI, Descartes retains the idea that the mental faculty of sensory perception is a passive faculty.
²See AT VIIIA.25: CSM I.210–11)
Moreover, when Descartes discusses ideas in *Mediation II*, he raises the question of material versus formal falsity in addition to the distinctions between innate, adventitious and invented ideas. In regard to the distinctions, Descartes explains the difference as

My understanding of what a thing is, what the truth is, and what thought is, seems to derive simply from my own nature, But my hearing a noise, as I do now, or seeing the sun, or feeling the fire, comes from things which are located outside of me, or so I have hitherto judged. Lastly, sirens, hippogriffs and the like are my own invention. (AT VII.38: CSM I.26)

Ideas defining a substance are innate, sensations are adventitious and imaginary constructs are invented ideas. Ideas of pure mathematics, the geometrical shapes, are innate, yet sensory ideas—the sounds, colors, etc., are adventitious. Of the adventitious ideas Descartes claims, “even if those ideas did come from things other than myself, it would not follow that they must resemble those thing.” (CSM II.27: AT VII.39)

Turning to sensory ideas and material falsity with some trepidation, Descartes notes that many sensory ideas are so unclear and indistinct that they are materially false. (CSM II.30: AT VII.44) Material falsity, as defined by Descartes, is falsity that occurs when an idea wrongly represents “non-things as things.” This is opposed to formal falsity, which he defines as error in judgment:

The chief and most common mistake is to be found here consists in my judging that the ideas which are in me resemble, or conform to, things located outside me. …if I considered just the ideas themselves as simply modes of my thought, without referring them to anything else, they could scarcely give me any *material* for error. (CSM II.26: AT VII.37)

Formal falsity, then, is an error in judgment about a substance(s) presented by an idea, whereas material falsity is an error within the idea itself—the error of idea phenomenally presenting a nonthing as though it were a thing. I will not here take the time to defend what is no doubt a controversial interpretation, for material falsity in Descartes’s work is a notoriously difficult notion. Nonetheless, following Wells’s interpretation, one might understand a materially false idea as an idea that fails to present or contain a genuine essence or principle attribute to the attending mind.¹ It is materially (metaphysically) flawed, as opposed to the epistemic flaw found in formally false ideas. To explain with Descartes’s example, if cold is nothing more than the absence of heat, then cold is not a genuine mode or attribute of a substance—therefore, the idea of cold does not represent or present any genuine essence to the attending mind. Sensory ideas are not usually sufficiently clear and distinct so as to view them as materially true ideas, *prima facie*.

In regard to the second difficulty, whether sensory ideas contain or present principle attributes or “true and immutable” natures to an attending, sensing mind, Descartes further muddies the water in *Meditation VI*. Descartes, after speculating that the faculty of imagination suggests that a body exists and that imaginary ideas of some geometrical figures are materially true (that is, sufficiently clear and distinct), notes that

But besides that corporeal nature which is the subject-matter of pure mathematics, there is much else that I habitually imagine, such as colours, sounds, tastes, pain and so on—though not so distinctly. Now I perceive these things much better by means of the senses, which is how, with the assistance of memory, they appear to have reached the imagination. (CSM II. 51: AT VII.74)

Here again Descartes separates the adventitious sensations or imaginings from the ideas of “pure mathematics.” Even so, after his proof that corporeal substances exist, Descartes claims that

What of the other aspects of corporeal things which are either particular (for example that the sun is of such and such a size or shape), or less clearly understood, such as light or sound or pain, and so on? Despite the high degree of doubt and uncertainty involved here, the very fact that God is not a deceiver, and the consequent impossibility of there being falsity in my opinions which cannot be corrected by some other faculty by God, offers me a sure hope that I can attain the truth even in these matters. (AT VII.80–1: CSM II.57)

This suggests that Descartes’s worry that sensory perceptions do not contain or present essences are resolved by his “God is Not a Deceiver” argument. Thus, even though sensory ideas are unclear and indistinct, they still may present a genuine essence to an attending mind.

V.

However, this leaves one with more questions than answers. Especially since Descartes then claims that:

I am also taught by nature that various other bodies exist in the vicinity of my body, and that some of these are to be sought out and others avoided. And from the fact that I perceive by my sense a great variety of colours, sounds, smells and tastes, as well as differences in heat, hardness and the like, I am correct in inferring that the bodies which are the source of these various sensory perceptions possess differences corresponding to them, though perhaps not resembling them. (AT VII.81: CSM I.56)
This suggests that Descartes separates himself a bit further from his Scholastic predecessors—at least more so than in his comments in *Dioptrics*. Claiming that sensory ideas correspond to, but do not resemble, the objects they represent to the attending mind suggest that for Descartes, these ideas do not contain or present the essences of that they represent, but instead cause the mind to think of the substances—to have a tertiary judgment that does contain the essence of external, sensed corporeal substances. Instead, second-grade sensory perceptions inform the mind of the external corporeal object. To explain, think of science fiction—a beep and red light emanating from a consul will inform a crew that the ship’s sensors detect “tacion” or some other fields off the bow of the ship. However, the red light and the beep, while corresponding to the presence of the “tacion” fields, do not resemble or contain the essence of the “tacion” field.¹

The trouble with this concept is that of reconciling it with material falsity. If Descartes denies that sensory ideas bear a representational, as well as a resemblance relation to the external corporeal substances they mean to represent and also holds that the ideas themselves do not contain or present essences to the mind, understanding how they do not present material for error is difficult. An attending mind with cognitive representations still retains the ability to transmit the essence or principal attribute to the attending mind. If one does not interpret Descartes as retaining the notion of representational phenomenal content, then one must create a more complex and perhaps far-fetched notion of material falsity.

Yet one more complication requires attention. In his *Replies*, Descartes indicates that the process of sensory perception does not only fail to transmit an essence from the external world to the mind, but also that sensory ideas such as colour and light are innate. Yet, one must take care in reading this passage, for Descartes also defines this sort of innate idea in a particularly non-Platonic manner:

Lastly, when we say that an idea is innate in us, we do not mean that it is always there before us. This would mean that no idea was innate. We simply mean that we have within ourselves the faculty of summoning up the idea. (CSM II.132: AT VII.189)

This allows one to understand that Descartes was not entirely “Platonic” in his understanding of how an idea, a sensory idea in particular, might represent or present an essence to an attending mind. If the physiological system is the creative cause of sensory ideas (*i.e.*, that which determines the content of a sensory idea), then the motions of animal spirits need not transmit and essence, but merely activate the passive faculty of sensation.

¹Alison Simmons, among others, suggests that for Descartes, human sensations are similar: sensations are teleological. Sensations alert us to that which is beneficial and harmful to us. God, on this interpretation of Descartes, gives us sensations to alert us to the world around us, allowing us to make judgments necessary to our survival. Just as the sensors on a space ship are purpose-build, to detect what is needed and what is to be avoided, so God created the human “sensors.” A. Simmons, “Sensible Ends: Latent Teleology in Descartes’ Account of Sensation.” *Journal of the History of Philosophy*, Vol. 39, Number 1, January 2001, pp. 49-75.
VI.

So, did Descartes reject Scholastic metaphysics entirely? It is not clear that he did. Clearly he left behind what he viewed as unnecessary and epistemically unjustified metaphysical entities such as sensorial and intentional essences. He also rejected the hylomorphic construct of the human in favor his dualism. Yet, the rejection of Scholastic mechanisms for sensation and imagination does not entail that he fully left the Aristo–Scholastic constructs of essences, attributes, causation and representational content with the unnecessary metaphysical “luggage.” Clearly, Descartes develops a unique take on an Aristotelian-style metaphysical world, removing that which he found unfounded and unnecessary and reconstructed what was left around a new method of justification and metaphysical picture of the human being.

Selected Bibliography


An Anthology of Philosophical Studies Volume 2


——. “*Descartes Embodied*.” Cambridge: Cambridge University Press, 2004


——. “Imagination, Phantasms, and the Making of Hobbesian and Cartesian Science,”


Part III

20th Century
1930 marked a turning point in Wittgenstein’s intellectual development. When he returned to Cambridge in that year, he thought, at last, that he discovered a method of doing philosophy that would allow him to treat all philosophical problems with professionalism and efficacy.

The nimbus of philosophy has been lost. For we now have a method of doing philosophy, and we can speak of skillful philosophers.1 Yes, I have reached a real resting place. I know that my method is right. My father was a businessman, and I am a businessman: I want my philosophy to be business-like, to get something done, to get something settled.2

There are two aspects to Wittgenstein’s method of deconstructing pseudo-philosophical problems that need to be distinguished: (1) describing actual linguistic practice, and (2) constructing hypothetical ‘language-games’. In a word, Wittgenstein’s method consisted in the investigation of uses of language, actual and possible. Both methods were, for Wittgenstein, an indispensable means of clarifying the ‘grammar’ of our words -- i.e., the appropriate contexts for using those words – and thereby dissolving pseudo-philosophical problems. Though (2) is often conflated with (1), it is important to recognize that it differs from it in important respects. (1) can be seen as functioning as a direct method of ‘proof’ (attempt to convince the reader of some thesis), and (2) as an indirect method of ‘proof’ -- proof by reductio. I try to clarify this distinction in what follows.

1Lectures (1930-32), p. 21.
Describing Actual Uses of Language

(1) involves describing and comparing concepts as they are used in their actual linguistic settings, noting that ‘Yes, this is what we say, and also this, and sometimes that...’, reminding ourselves of uses of language that we have forgotten or overlooked. By reminding the philosopher of the role and function of expressions of our language, Wittgenstein hoped to weaken the hold that misleading analogies have on him. For many of the ‘problems’ that arise in philosophy, Wittgenstein recognized, rest on overstretched analogies, which the philosopher can dissolve by reminding him/her-self of how certain expressions are employed in their actual contexts of usage.

… the sentence only seems queer when one imagines a different language-game for it from the one in which we actually use it. [PI §195]

Describing the various ways we actually use words helps relieve us from a strained way of looking at them. This is the point of Wittgenstein’s remark that:

A human being is imprisoned in a room, if the door is unlocked but opens inward; he, however, never gets the idea of pulling instead of pushing against it. [RFM III, 37]

This is a recurring image (imprisonment) Wittgenstein had of the philosopher, who he felt he must rescue. Pushing against the door is like the philosopher’s treating philosophical problems like scientific problems: ‘We still don’t know that; but it is knowable and it is only a matter of time before we get to know it!’ [CV p. 40] For Wittgenstein, this is the wrong attitude to take to philosophical problems:

It was true to say that our considerations could not be scientific ones. … And we may not advance any kind of theory. There must not be anything hypothetical in our considerations. We must do away with all explanation, and description alone must take its place. And this description gets its light, that is to say its purpose, from the philosophical problems. These are, of course, not empirical problems; they are solved, rather, by looking into the workings of our language, and that in such a way as to make us recognize those workings: in despite of an urge to misunderstand them. The problems are solved, not by giving new information, but by arranging what we have always known. [PI §109]

Philosophy simply puts everything before us and neither explains nor deduces anything. – Since everything lies open to view there is nothing to explain. For what is hidden, for example, is of no interest to us. [PI §126]
It is important not to misunderstand what Wittgenstein is saying here, as commentators often have. There is a way in which I think he would grant that philosophy *is* a science, and *does* construct theories. One way to think of a scientific theory is as a statement that organizes a set of concepts in a meaningful way by explaining the relationship among them. The various sciences assume there is some underlying order in the universe; that events, whether they involve molecules or human beings, are not haphazard. They follow a pattern that is sufficiently regular for generalizations to be made about them. Generalizations are crucial to science because they place isolated, seemingly meaningless events in patterns we can understand. It then becomes possible to *explain* why something happens and to *predict* that it will happen again under the same conditions. In addition to the natural sciences, which try to explain *physical and biological phenomena* in this respect, and the social sciences (sociology, economics, psychology, political science, and anthropology) which try to explain *human behavior*, there is a place (as I think Wittgenstein would grant) for a science that makes *language* its object of study. This science would place seemingly meaningless phenomena regarding linguistic practice in a general framework that would enable us to *explain* why philosophical puzzlement/conceptual tensions occur, and enable us to *predict* that they will recur again under certain conditions (ways of interpreting language). I don’t think taking philosophy to be a ‘science’ in this respect (a science of the misuses of language that lead to conceptual muddles) would have been unwelcome to Wittgenstein, as passages like the following suggest:

Language is an immense network of easily accessible wrong turnings. And so we watch one man after another walking down the same paths and we know in advance where he will branch off, where walk straight on without noticing the side turnings, etc. etc. What I have to do then is erect signposts at all the junctions where there are wrong turnings so as to help people past the danger spots. [CV p. 18]

… it isn’t as though everything we say has a conscious purpose; our tongues just keep going. Our thoughts run in established *routines*, we pass automatically from one thought to another according to the techniques we have learned. And now comes the time for us to survey what we have said. We have made a whole lot of movements that do not further our purpose, or that even impede it, and now we have to clarify our thought processes philosophically. [CV p. 64]

As long as there continues to be a verb ‘to be’ that looks as if it functions in the same way as ‘to eat’ and ‘to drink’, as long as we still have the adjectives ‘identical’, ‘true’, ‘false’, ‘possible’, as long as we continue to talk of a river of time, of an expanse of space, etc. etc., people will keep stumbling over the same puzzling difficulties and find themselves staring at something which no explanation seems capable of clearing up. [CV p. 15]

He [Wittgenstein] held that though the ‘new subject’ must say a great deal about language, it was only necessary for it to deal with those
points about language which have led, or are likely to lead, to definite philosophical puzzles or errors.¹

The reason we ‘know in advance’ where ‘someone will walk straight without noticing the side turnings’ is because we have observed this phenomena repeatedly. Observed ‘routines’ enable us to frame generalizations and predictions. Just like the child who characteristically overgeneralizes and says ‘sleeped’ instead of ‘slept’, ‘brang’ instead of ‘brought’², philosophers too are bound to ‘stumble over the same puzzling difficulties’ by overstretching analogies. It is part of the philosopher’s task to bring to conscious awareness the unacknowledged misuses of language (overstretched analogies) that lie behind these puzzling difficulties.

In the respects I have outlined, I don’t think it would be wrongheaded to conclude that Wittgenstein was in the business of constructing ‘theories’. Like the good scientist, Wittgenstein collected a seemingly endless stream of empirical evidence (drawn from actual linguistic practice), which served as a basis for generalizations (about how certain features of language can mislead us), explanations (of how pseudo-problems arise), and predictions (about how they are bound to rise again and again, given certain features of language). A science of this sort is surely one that is desperately needed (efforts to understand the ‘human condition/predicament’, to borrow some post-modern jargon, demand it), and we can usefully see Wittgenstein’s philosophical method, applied to diverse regions of language, as inviting its establishment. If there is anything that unifies Wittgenstein’s philosophical explorations in these disparate linguistic settings, it is his desire to get us to see how certain (deeply entrenched) features of language can lead us (and have led many philosophers) astray.

It is a great irony that Wittgenstein’s philosophical method, which has been condemned by many of his fiercest critics as being the most belligerent toward science, adheres more strictly to the canons of science as I have described them above than the approach of many of the staunchest defenders of the use science in philosophy! One factor that obstructs this point from view is that Wittgenstein was mostly preoccupied with gathering empirical data, leaving the task of organizing it, and drawing the proper morals (generalizations, predictions) to the reader. As Wittgenstein cryptically puts this point:

> If this stone won’t budge at present and is wedged in, move some of the other stones round it first. -- All we want to do is straighten you up on the track if your coach is crooked on the rails. Driving it afterwards is something we shall leave to you. [CV p. 39]

What distinguished Wittgenstein’s ‘theories’ from other philosophical theories (like platonism in the philosophy of math, or psychologism in the philosophy of mind) was that their aim was not to solve pseudo-problems, generated by overstretched analogies, but to dissolve them; to explain them

¹Moore (1955) p. 27.
²These examples are offered by Chomsky (1986) p. 227.
away; to show why they are illegitimate. For Wittgenstein, there’s nothing wrong with philosophical theorizing per se; there is something wrong with philosophical theorizing designed to solve pseudo-problems, as science does scientific problems. As he says, e.g.

It is the business of philosophy, not to resolve a contradiction by means of a mathematical or logico-mathematical discovery, but to make it possible for us to get a clear view of the state of mathematics that troubles us: the state of affairs before the contradiction is resolved. (And this does not mean that one is sidestepping a difficulty). [PI §125]

We can sum up Wittgenstein’s advice to the reader as: ‘Don’t wed yourself to a “philosophical problem” too quickly! Don’t become too dazzled by it! It might be a snare!’

Since pseudo-problems are dissolved and not solved, no new knowledge is gained. As Moore recounts of Wittgenstein:

He also said that he was not trying to teach us any new facts: that he would only tell us trivial things – ‘things which we all know already’; but that the difficult thing was to get a synopsis of these trivialities, and that our ‘intellectual discomfort’ can only be removed by a synopsis of many trivialities – that ‘if we leave out any, we still have the feeling that something is wrong.’ In this connection he said it was misleading to say that what we wanted was an ‘analysis’, since in science to ‘analyze’ water means to discover some new fact about it, e.g. that it is composed of oxygen and hydrogen, whereas in philosophy ‘we know at the start all the facts we need to know.’

The goal of ‘analysis’ in philosophy, as Wittgenstein conceived it, is to clarify the ‘grammar’ of our words, i.e., the appropriate contexts for using those words. This involves describing what we already know but have forgotten. There is of course a sense in which ‘analyses’ of this sort are hidden from view (and so we might think that analyzing here means finding something new), but that is only because we overlook the familiar aspects of word-usage, and need to be reminded of them.

The aspects of things that are most important for us are hidden because of their simplicity and familiarity. (One is unable to notice something – because it is always before one’s eyes.) [PI §129]

How hard it is to see what is right in front of my eyes! [CV p. 39]

---

1Moore (1955), p. 27.
Exploring Imaginary Uses of Language


I shall in the future again and again draw your attention to what I shall call language-games. These are ways of using signs simpler than those in which we use the signs of our highly complicated everyday language. Language-games are the forms of language with which a child begins to make use of words. The study of language-games is the study of primitive forms of language or primitive languages... When we look at such simple forms of language the mental mist which seems to enshroud our ordinary use of language disappears. We see activities, reactions, which are clear-cut and transparent. [BB p. 17]

The notion of a ‘language-game’, like that of a ‘conceptual framework’ or ‘discourse’ or ‘practice’, is a loose concept: it has ‘no precise definition or decisive and non-arbitrary criterion of individuation’. Wittgenstein uses the notion of a ‘language-game’ to refer not only to actual uses of language – which might involve very basic ‘moves’, as in teaching language to children, or more sophisticated/complex ‘moves’, as we find in everyday discourse – but also to hypothetical or invented uses of language, which again might be basic or complex. 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]

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. This is a subtler way of deconstructing pseudo-problems than the method discussed in the preceding section – of describing actual uses of language -- since it operates in an indirect manner. Hypothetical language games involve abstracting concepts from their normal circumstances of application, as in the following diagram.

---

1Barnett’s expression in (1990) p. 49.
'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]

But how does it ‘disperse the fog’ to take such departures from reality? How do language-games help us see things from a ‘completely new angle’?

In answering this, I have found it useful to draw an analogy with surrealism in art. 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. An example to illustrate.1

---

1© Chrysoula Gitsoulis
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]

Unlike most surrealist works, however, Wittgenstein’s re-arrangements of language are not arbitrary, but strategically chosen. Pseudo-philosophical problems, remember, are generated by overstretching the ‘limits of language’: interpreting the sense and function of words/sentences in one use-context/‘language-game’ in terms of their sense/function in another use-context/‘language-game’. It is in this respect that we overstretch the ‘limits of language’, thereby generating false interpretations that lead to pseudo-philosophical problems. In a hypothetical language-game we are invited to imagine a world where the false interpretation is correct. What else would have to be true if it is correct? A hypothetical language-game fills in the details. The idea is that by accumulating enough of these details, we might at last come to see that we have been in the grip of a false interpretation. So hypothetical language-games, we can say, function like reductio arguments.

---

1 Or 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’.
This is why this aspect of Wittgenstein’s method is more subtle than that of describing actual uses of language: it is an indirect method of ‘proof’ -- of getting us to see our mistaken assumptions.

An example to illustrate. In the note to PI §151, Wittgenstein tries to clarify the grammar of our concept of understanding. Can it be correct to think of understanding as a ‘mental state’? Well, suppose it is correct. What else would have to be true? To answer this, we need to consider how we actually employ the concept of a ‘mental state’. We apply it to experiences such as depression, excitement, and pain, among others. And what is true about how we apply these terms? We say, e.g., as Wittgenstein notes:

**Actual Use-context**

- ‘He was depressed the whole day.’
- ‘He was in great excitement the whole day.’
- ‘He has been in continuous pain since yesterday.’
- ‘When did your pains get less?’

But can we speak of ‘understanding’ in these ways? To answer this, we need to construct imaginary use-contexts where the term *is* used in these ways. Again, using Wittgenstein’s examples:

**Imaginary Use-context**

‘Since yesterday, I have understood this word.’ ‘Continuously’ though?
‘When did you stop understanding the word?’

As we can see, it is ‘ungrammatical’ to apply certain *temporal* concepts to ‘understanding’ in the way we do to mental states: we speak of being in continuous pain, or of a pain being interrupted for several minutes, or of suddenly ceasing to feel pain, but it is grammatically jarring to regard understanding as ‘clockable’\(^1\) in this way. Moreover, it is grammatically jarring to apply concepts of *intensity* to understanding in the way we do to mental states: we describe a pain as intense, or an emotion as strong, but it is ‘ungrammatical’ to describe understanding in these terms.\(^2\) One might do so, but then he is using the term in nonstandard or conventionally unacceptable ways. Similarly, one might use a kitchen table as a TV stand, but this is not conventionally acceptable, and to bring this out to someone who decides to use it in this way, you might try placing a checkered table cloth over the table, or kitchen chairs around it, and a napkin holder next to the TV (you would be creating an imaginary language-game). If this provokes laughter, you have made your point, though indirectly of course. It is similar with words.

\(^1\)McGinn’s expression in (1984) p. 5.
\(^2\)We can also see Wittgenstein’s so-called ‘private language argument’ as taking the form of a reductio.
Laughter might be the appropriate response to someone who uses words as in the imaginary use-context above (‘I stopped understanding the word yesterday’, etc.). Laughter is a sign that an interpretation is out of place, as in the joke:

Patient: I broke my arm in two places.
Doctor: Don’t go to those places.¹

This is why Wittgenstein makes use of jokes to illuminate concepts. They are supposed to help us recognize that we are in the grip of a false interpretation.

Rearrangement of familiar facts in Wittgenstein’s hypothetical language-games is another reason (in addition to those mentioned in §3.1) Wittgenstein’s writing gives rise to a feeling of eeriness (a bit like the feeling we have as tourists traveling to unfamiliar lands). Witness for example the following language-games:

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.

The eerie (or comic) aspect arises from using familiar words in unfamiliar ways, just as in surrealism it arises from rearranging elements of familiar settings. 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 with our actual linguistic practices. In this way we might see (the contrast might help bring to light) that communication does not demand that every word must have something for which it stands -- that something being its meaning -- as the Augustinian picture (a pseudo-picture) assumes.

What we have overlooked, what we have forgotten, for Wittgenstein, is often what is right before us: how language actually functions. It is to this

¹This joke is from Richard Gilmore (1999) p. 96.
Wittgenstein’s Treatment of Philosophical Problems

world (for Wittgenstein both the source of and final court of appeal for philosophical disputes) that Wittgenstein was constantly drawing our attention (either directly or indirectly), by making the insignificant significant, the ordinary extraordinary. It is perhaps in this respect more than any other – in recognizing the elementary sources of confusion that lie at the root of many of our seemingly most ‘profound’ philosophical problems -- that his work was revolutionary; and therein, I think, lies his most valuable contribution to philosophy.

Bibliography

Butler’s skepticism toward a pregiven subject-position, whether it invokes the paternal law or its never-ending displacement, is very much indebted to Foucault’s method of genealogy; that is to say, ‘a form of history which can account for the constitution of knowledges, discourses, domains of objects etc., without having to make reference to a subject which is either transcendental in relation to the field of events or runs in its empty sameness throughout the course of history’ (Foucault, 1980: 117).

The subject does not stand before or after the law, but is inaugurated by the very relations of power which it may seek to undermine. While the law of the subject is an ideal construct, inseparable from its materialized effects, it is not a simple cause; it does not relate to concrete manifestations of power within a linear trajectory. If it is true, for example, that the assumption of sex is determined and constructed according to the repeated enactment of certain dominant values – and not only the repetition of norms but likewise the innumerable modes and elaborations which enter into any concrete regulative practice –¹ it is no less true that those values are given life through the assumptions and reiterations that they occasion (Butler, 1993: 14). There are no unilateral powers or commands: the historical instantiations of a male-identified sexuality which covers over its own exclusions, its own abject projections, are open to change and reevaluation precisely because they are constituted through time (Butler, 1990: 40).

¹The point being that the quality of an action is not solely determined by its conformity to law: we must also consider how the subject complies with such restraints, how he justifies them and puts them into practice (Foucault, 1990b: 26-30).
The rearticulation of norms is, accordingly, a historical practice enabled by those uninhabitable zones of difference that have been disavowed in the contemporary field of power. An absolute outside does not provide us with the means to either rethink or reconfigure heterosexist principles of repudiation: resistance arises from the living structures of language which are themselves coextensive with power. Butler will therefore make the argument, with Lacan and Kristeva in mind, that playful resignification need not be construed as purely imaginary: ‘Importantly, the erotic redeployment of prohibitions and the production of new cultural forms for sexuality is not a transient affair within an imaginary domain that will inevitably evaporate under the prohibitive force of the symbolic’ (1993: 110). Unless the imaginary and the symbolic are deeply interwoven, the effects of “heterogeneous drives” are ineffectual. Those drives, for Butler, cannot preexist the symbolic saturation of subjectivity which they oppose and contest. It is for this reason that the futural imaginings of contestation are dependent upon the concrete practices of subversion, even if the results of subversion, due to the social temporality of discourse, are provisional. It also for this reason, despite Foucault’s influence on gender theory, that Butler is critical of those moments in his writing in which he appeals to an ontological form of resistance set apart, ostensibly at least, from historical modes of contestation. The reification of sexual identity, as a gendered essence immune to social and political resignification, thereby denies the very possibility of deconstructing oppressive subject-positions. In Foucault’s defense, I will show that his appeal to ontological invariants is neither an accident nor a simple contradiction in his thinking. A more sensitive reading of his oeuvre is required, one which takes into consideration Bataille’s early and lasting influence. Moreover, this reading will point us toward a more complex understanding of Foucault’s politics of sex and gender, by expanding his notion of radical resistance to include not only that which is politically and historically specified, i.e., postmodern subversion, but also that mode of transgression which disrupts the very proliferation of techniques and powers that today mark us as administered subjects.

In the first volume to Foucault’s History of Sexuality it is clear that an analytics of power has replaced another representation of power which he terms ‘juridico-discursive’ (1990a: 82/1977a: 109). No longer conceived as repressive or negative, power is now described as a ‘moving substrate of force relations’ (1990a: 93/1977a: 122). The methodological shift, needless to say, has had a profound impact: Elizabeth Grosz writes that one of Foucault’s major innovations, namely, the rethinking of sex as an effect as opposed to a ground, has contributed to the deconstruction of sex/gender binaries (1994: 212). When David M. Halperin puts forth that the invention of homosexuality accompanied the eighteenth-century discovery and definition of sexuality in general, as well

---

1For Butler, these zones of difference are not entirely invisible: they are constructed within a narrative of exclusion which identifies them as incoherent.
2Furthermore, whether maternal or polymorphous, these drives will facilitate the workings of power insofar as they disguise the cause of their production. They are a manifestation of power which diverts attention away from its historical sedimentation.
3They must be provisional since the effects of discourse, whether subversive or reactionary, always exceed our stated intentions and explicit purposes (Butler, 1992: 10).
as the nineteenth-century interpretation of sexuality as an underlying behavioral principle, there is an unmistakable Foucaultian strain: ‘Sexuality, on this latter interpretation, turns out to be something more than an endogenous principle of motivation outwardly expressed by the performance of sexual acts; it is a mute power subtly and deviously at work throughout a wide range of human behaviors, attitudes, tastes, choices, gestures, styles, pursuits, judgments, and utterances. Sexuality is thus the inmost part of an individual human nature’ (1990: 26). Moira Gatens explicitly draws from the genealogical approach as she challenges the masculine construction of female attitudes, difference, sexuality, and politics (1991: 70). Yet Foucault himself postulates a realm of radical resistance apart from sex and power: ‘The rallying point for the counterattack against the deployment of sexuality ought not to be sex-desire, but bodies and pleasures’ (1990a: 157/1977a: 208). De Lauretis accounts for the discrepancy by gesturing toward an irresolvable tension between affirmative political action and its theoretical negation: the counterattack is in this way a manifestation of power that has yet to be deconstructed (1997: 268). In spite of his own arguments, Foucault seems to indicate that there are certain emancipatory activities which cannot be reduced to the space or logic of sexuality. This in turn presumes a distinction between history and sex which reinvests the body as a precultural phenomenon: ‘In a sense, for Foucault, as for Nietzsche, cultural values emerge as the result of an inscription on the body, understood as a medium, indeed, a blank page’ (Butler, 1990: 66). Consequently, Foucault has been accused of paradoxical conservatism: the feminist text, according to Grosz, should uncover the dominant normative ideals in a phallocentric environment (1994: 22). It is also imperative to open up the production of new possibilities, investments, and subject-positions that are not currently sanctioned by the masculinist regimes of power (1994: 23). By having recourse to a strategy of revolt that places itself beyond the range of historical analytics, Foucault returns to a notion of self-identity which is ahistorical and essentialist. What seems to be implied, by neutralizing sexual difference and degendering political critique, is a foreclosed limit to the deconstruction of patriarchal discourses. A frequent response to this tactical lapse is to fold the method back upon itself: ‘Bodies and pleasures cannot be understood as fixed or biologically given constants, somehow outside of or beyond the constraints of power—no matter how much Foucault himself may have yearned for a disinvested ground, a pure datum, onto which the operations of power can be directed, no matter how much he believed, in spite of himself, in harmless, timeless pleasures and bodies as yet unmarked by power’ (Grosz, 1994: 218).

The same tensions which have been detected in Foucault’s later texts are also to be found in his ‘Preface to Transgression’ which distinguished, by keeping with Bataille, eroticism and sexuality (1977b: 33). He does not write that sexuality can be liberated from the effects of language, or that it brings us closer to the realization of our deepest secrets, but only that it returns us to a limit which can not be crossed. There is no freedom apart from the consciousness of a limit: eroticism is itself the experience of sexuality which destroys itself in a language known only by those who speak (Foucault, 1977b:
The historicization of sexuality is therefore implied by its transgression: the limit and its impossible crossing form a single point of convergence. And yet here, as elsewhere, Foucault alludes to a profanation that is empty, without object or content: ‘In that zone which our culture affords for our gestures and speech, transgression prescribes not only the sole manner of discovering the sacred and its unmediated substance, but also a way of recomposing its empty form, its absence, through which it becomes all the more scintillating’ (1977b: 30). When the whole essay is taken into consideration, especially in relation to Bataille, it appears that transgression is the opening up of substance to its own exuberance. This has to be carefully understood: profanation without the form of substance is not equivalent to the assumption or identification of substance as pure negation: ‘Transgression contains nothing negative, but affirms limited being—affirms the limitlessness into which it leaps as it opens this zone to existence for the first time’ (Foucault, 1977b: 35). What might be called the inexpressible or unmediated aspect of profanation is simply the affirmation of substance to the point of its disappearance. And to the extent that the normalization of society is the historical result of a life-administering power (Foucault, 1977a: 144), we should not be surprised if the affirmation of substance includes the mortal matter of specific bodies and pleasures.¹

That there is an unconditional element, however, cannot be disputed. While the history of sex and sexuality emerges within a configuration of power that stabilizes identity across time, through the materialization of time, there is an element of transgression which cannot be reduced to social temporality. What might be called either substance or energy, in the works of Nietzsche, Bataille, and Foucault, has been perceived as a vitalistic return to Hegel’s notion of desire in the life-and-death struggle. Hence, for Butler, questions of revolt are still situated within a matrix of power that divides itself according to life-affirming and life-negating forces: ‘[I]n prescribing the overthrow of juridical models of power, Foucault faults such models for subduing the life-affirming energy characteristic of productive power.... Either as pure energy, the will-to-power, or life itself, productive desire seems less an historically determined than an occasioned desire which, in its origins, is an ontological invariant of human life’ (1987: 227-28).² It would appear to be the case that his ontologization of desire is linked to the fixed constants of sexual identity which Foucault otherwise rejects. But this supposes two things which have yet to be proven: 1) that the metaphysics of substance implies an immutable ontology; and 2) yielding to an unconditional or ahistorical element is equivalent to taking up an unconditional or ahistorical position. Looking back to Bataille we see that these two propositions are not always valid: it is quite conceivable that

¹In both the ‘Preface’ and History of Sexuality vol. 1, Foucault links the disappearance of the subject with transgression (pp. 34 and 138 respectively), which seems to confirm the idea that revolt is based less on resignifying practices (even as they should never be underestimated) than the limit of power which is neither limited nor limitless but marks us nonetheless as an extreme passage.

²These remarks, it should be noted, are located in part four which retroactively supplemented the dissertation proper.
an unmediated substance is neither an unchanging materiality nor a self-grounded locus of agency.

Although being cannot be represented to us outside of words, outside of the constructions and thematizations of meaning, it is the substance of all things, linguistic or otherwise, which can not be fixed within a finite circulation of symbols and signs. It is surely appropriate to interrogate the theoretical moves which establish universal, indisputable values beyond the scope of power. At the same time, however, we should be aware that chance permutations of substance need not be confined to human parameters: ‘What’s called substance is just a provisional equilibrium between the spending (loss) and the accumulation of force. Stability can never exceed this short-lived, relative equilibrium; to my mind, it’s not and can’t ever be static’ (Bataille, 1988: 15). Substance, in other words, is not a value. It is not a thing. It is not an object in itself calling for us to understand it. We only know of substance because we exist in a world which surpasses us. It can be said that we understand the world which surpasses us to the extent that we come into contact with it, but it does not follow from this that we have penetrated its secrets—for there are no secrets. Substance eludes us, not because it is out of reach, but only in the sense that it is itself a provisional equilibrium. Bataille’s theory of expenditure, it will be claimed even by those who admire him, necessitates an absolute loss of self in the substance of self which is pure, unconditional heterogeneity: ‘In order to escape the systematization and domination of rationality, the self must be fully lost in immanence in order to establish the whole man [l’homme entier] as a completely unsubordinated realm.... And communication is no longer a teleological activity, but rather a function of both love and evil where the relation between two beings is put entirely at risk, exposed to the vagaries of chance’ (Weiss, 1986: 137-38). Bataille himself does not imagine that we can relate ourselves to immanence in a completely unsubordinated realm, for this implies a realization of self that is no longer bound to the material conditions of life. If we are affected by chance it is not in the sense that we are able to take up a prediscursive position in respect to it, but rather in the sense that we are affected by a movement of substance and energy that has no bounds: ‘In this world there is no immense undertaking that has any other end than a definitive loss in the futile moment. Just as the world of things is nothing in the superfluous universe where it is dissolved, the mass of efforts is nothing next to the futility of a single moment’ (Bataille, 1989: 102-03).

This view of chance, aligned with its ontological exteriority, is easily distinguished from Butler’s. While it is certainly not the case that the author of Bodies that Matter reduces material entities to an idealistic form of language, as if language were not itself a materialized process of signification, she leaves no doubt that every articulation of matter is bound to its referent as a mutually contaminated realm of signification (Butler, 1993: 68-9). Butler does not deny that we are mortal creatures, that we exist in a world of pleasure and suffering, or that the body forms a troubling blind spot to the intelligibility of speech (1997: 11 and 155). What she does argue is that the contingency of discourse can be strategically used to undermine the reified positionalities of gender that have been taken up in that discourse. Generally speaking, the gaps and fissures
that are associated with discursive enactments are what allow for subversive transformations of language. This includes language which has proven to be extremely hurtful or offensive: ‘Those who seek to fix with certainty the link between certain speech acts and their injurious effects will surely lament the open temporality of the speech act. Such a loosening of the link between act and injury, however, opens up the possibility for a counter-speech, a kind of talking back, that would be foreclosed by the tightening of that link’ (Butler, 1997: 15). The meaning of speech is equivocal. It is open to a multitude of contradictory appropriations that can be turned against one another, and the subject which bears this equivocal meaning is a moving manifold of signification. Defined by a logic of exclusion, the normalized subject establishes an autonomous realm of truth and knowledge that is set apart from its abjected other. But at the same time, the socialized relations between self and other are precarious: the founding acts of repudiation that help to solidify mainstream, heterosexist categories of subjectivity are haunted by the very instabilities which these acts inadvertently produce.1

For Butler, then, the presentist view of the subject is undermined through an unexpected redeployment of its terms: ‘The replication of heterosexual constructs in non-heterosexual frames brings into relief the utterly constructed status of the so-called heterosexual original. Thus, gay is to straight not as copy is to original, but, rather, as copy is to copy’ (1990: 41). Whether normative or marginalized, every marked position, every formation of self-identity, is subject to further elaborations of meaning. The discursive field of sex and power is illimitable. But if we look to Bataille’s general economy we shall find yet another critique of presence based upon the illimitable historicity of language. The subject, for Bataille, is confronted with its unnecessary existence when its conditional reality has been affirmed as unconditional. If Butler is correct that losing one’s place in language is the result of being exposed to an unknown future, to the unpredictable consequences of an injurious address, then it seems to follow that Bataille’s sacrifice of language to an absolute outside, to the unconditional absence of ourselves, is even more vertiginous. It would thus be incorrect to assume that transgression, in Bataille’s use of the word, is devoid of political relevance. Likewise, Foucault’s strategic recourse to ‘bodies and pleasures’ should not be viewed as an unseemly aberration, as if the subjectification of the self were completely overturned in a naïve rush to embrace the inner truth of humanity. Contrary to several influential readings of Foucault, it seems extraordinarily unlikely that he forgot so much of his own work, so much of his analytics and discourse of power, that he would in the last instance retrace his steps and return us to a privileged view of absolute freedom.

The question of resistance vis-à-vis modern formations of power is perhaps the most debated question amongst Foucault scholars. One line of criticism, put forth from a humanist perspective, challenges whether there is in fact an ethical mode of resistance to be found in Foucault’s works. Nancy Fraser, drawing heavily from Habermas, wonders how it is that removing our ethical and political foundations will ever allow us to critique oppressive social structures.

---

1Hence, they are not truly founding acts; they are echoes and traces of repeated citations.
Foucault, it would seem, needs to offer us an alternative to humanist accounts of individual and political freedom, but this he never does. She argues that he instead makes a vague claim that ‘protest urged in the name of the pleasures of our bodies may have greater emancipatory potential than that made in the name of the ideal of autonomy’ (Fraser, 1996: 33). Foucault’s defenders, often inspired by his turn to ancient Greek ethics, remind us that his deconstructive interrogations of the self can and should anticipate new forms of radical subjectivity. Freedom is thereby claimed to be ethical in the sense of being a continual practice of subversive, activist, and artistic self-formation.¹ But what is too often missed, by opponents and defenders alike, is the contingency of discourse exposed by Foucault’s genealogical method—not only of this or that discourse, but of discourse itself. The fallacy shared by his humanist critics and his postmodern disciples is that in both cases the subject is entirely constituted by power and discourse: either it is guided by a form of power which regulates itself according to principles of liberation and autonomy, or it is motivated by a constant reworking of its own performative boundaries. But in either case discourse is absolutely paramount.

If we look back, as I have tried to do in this essay, to Foucault’s theoretical connection with Bataille, it becomes clear that the ubiquity of discourse is no less critiqued and challenged by the former than the latter. The rallying point against the deployment of sexuality cannot be situated only within the matrix of sex-desire because that matrix is itself discursive. Bodies and pleasures do not bring us closer to an absolute, utopian kind of freedom, but they do in fact bring us closer to a limit—the limit of subjectivity and life. This explains to us why Butler is wrong to presume that Foucault is returning us to a vitalistic principle: if that were true, then he would not gesture toward an element of resistance that defies the very administration of life and sexuality which pervades our modern age when he writes that ‘death is power’s limit, the moment that escapes it’ (Foucault, 1990a: 138/1977a: 182). Most commentators agree that Foucault breaks from his earlier metaphysical ‘errors’ when he writes *Discipline and Punish*; yet in this very book in which he describes modern punishment as an entirely mediated process, he not only looks to subversion (i.e., that which serves as a strategic reversal) as a principle of revolt, but perhaps more importantly to that which counters strategy and discipline altogether. Just as he does in the first volume of his *History of Sexuality*, Foucault intimates toward the final pages of *Discipline and Punish* that it is utility itself (encompassing life and power) which has become one of the most pressing modern political issues (1995: 306/1975: 358). If this were a masculinist return to pure energy and life-affirmation, a return that could be rightfully associated with male/female dichotomies of activity/passivity, then the revolt against power would not be self-sacrificing. And lastly, though it has been said by some that there is no revolt against bio-power that can be sustained for too long via self-sacrifice (Simons, 2001: 85), once again we

¹According to Jon Simons, and many others, freedom is thus distinguished from oppression in terms of whether or not power is opened up to a continual reevaluation of its own configurations. This is what Simons identifies as the affirmation of agonism (2001: 86).
should keep in mind Foucault’s link to Bataille. If the former does not fully, explicitly theorize his recourse to a metaphysical form of transgression in his later works, perhaps it is because he has already done so in his earlier essay on the latter, in which we have already learned that a self-sacrificing revolt does not always place us on the other side of life. What it must do, however, is show us the limits of our own self-identity—our own manipulated, constructed, totalized, and ritualistically performed self-identity. Moments of uselessness, whereby we release ourselves from actions and attachments otherwise determined by strategy and success, need not push us beyond this life to remind us that human life itself is never fully mastered.

Bibliography


From where does our wrath towards others spring? What is the psychical
well from which totalitarian violence, as seen in the past century,
emerge? Does that even matter politically? Further, what are the
structural frameworks that contribute to the menace of totalitarian
violence? Of course, these questions have, in large measure, dominated the
philosophical landscape since the Holocaust. Many contemporary continental
philosophers such as Derrida and Blanchot have argued that the events that
occurred during the Holocaust must remain fundamentally opaque and
incomprehensible. Two students of Heidegger, Levinas and Arendt, have
offered dramatically different visions for what was at the root of this
violence, of what the conceptual framework was at stake, and of what could
quite possibly preclude this new incarnation of aggression towards others by
means of a non-totalitarian framework. In this essay, I will, first, briefly inquire
into the notions of violence as delineated by Heidegger and Levinas. Only then
will it become clear why Arendt’s conception of the impetus towards
violence—latent within her work—surpasses the first two. Later, I will
consider how these notions of violence are related to ‘liberalism’ and
‘modernity’. In this light, we will I consider why an anti-liberal bulwark
against totalitarian violence best understands what is at stake in the modern
world, and what is utterly unique about said world. As a result, I am able to
show why the categories of ‘conscience’ and ‘action’ are at the heart of the
debate.
Heidegger and Levinas on ‘Violence’

For Heidegger, we already know our existence, but in order to preclude an impoverished return to the circle of interpretation, we must do violence to the everyday perspective (the perspective of the they-self\(^1\)), so as to become more familiar with its structure, thereby making existence more ripe with meaning and significance. This, therefore, reveals that the meaning of our existence lies within our existence, not outside of it. Without this violence, the everyday interpretation of existence (das man) cannot be heroically transcended (although, it is never altogether deserted), and subsequently thematized.

For Levinas, the originary ethical moment lies in the face-to-face encounter with the Other, where the powers and sovereignty of the self are called into question. Although he calls it a ‘pacific encounter’ with the always radically separate Other, Levinas’ face-to-face revelation offers a direct challenge to one’s imperious egoism, and constitutes ethics as such, whereby our freedom offers us the options of violence and detestation towards the Other or responsibility, obligation, and respect. The ethical encounter with the Other is the very foundation, then, of our freedom.

The imperceptible face of the Other, for Levinas, speaks, and, upon its epiphanous explosion into our world, it first commands us not to commit murder. Levinas claims that, ‘this infinity, stronger than murder, already resists us in his face, is his face, is the primordial expression, and is the first word: “you shall not commit murder.”’\(^2\) This commandment (from which all others derive), however, is not compelling or of divine authority, as it is not, claims Levinas, founded, and, therefore, has no means of making me obey.

With regard to what Levinas calls the ‘ethical resistance’, an insurmountable resistance\(^3\) not measured in force qua force, but rather in the realization that the Other resists my attempts at comprehension, domination, and empire, the only manner in which the self can suppress the Other is to annihilate it completely. And it is with this conception of annihilation, then, that Levinas tells us that, although we can never actually do so, we desire to kill the Other (the only being I, indeed, wish to kill), as it is altogether beyond the sphere or scope of my powers of comprehension, relation, and domination.

A More Complete Picture of Violence

Let’s consider these two accounts of violence, Levinas’ account and Heidegger’s, from the point of view of intent. As Arendt rightly points out, in

---

\(^1\)The ‘they-self’ is what Heidegger calls human beings (what Heidegger famously refers to as ‘dasein’) who, in their “fallenness”, become immersed in the objects of average-everydayness (gossip, idle chatter, curiosity, and the like), and, thus, lose their own-self (eigentlichkeit). A violence towards the everyday perspective, would entail an embracing of the angst that accompanies the recognition of our own finitude, and that our choices, projects, and possibilities matter.


\(^3\)Ibid. 197.
Levinas and Arendt on Violence and Conscience: The Disharmony of Everyday Life

our world of liberal political jurisprudence, this is the almost invariant perspective we adopt toward violence: for example, we consider an act of killing to be murder only insofar as it involves intent. This, Arendt calls with derision, the ‘subjective factor’.  

Surely, it is not only a Levinasian, ‘sadistic’ impulse towards that which is utterly alien and beyond my sovereign powers of comprehension which fuels the descent into totalitarian violence, but it is also what we might call a ‘masochistic’ one. For, we clearly also commit violence in the name of that which reminds us most profoundly of what we find contemptuous about ourselves, and, in doing violence towards the Other, I foster for myself a form of self-affirmation that I am not that.

So, it is also those who remind us of what we detest in ourselves the most, which provokes in us a desire for their annihilation, as we no longer wish to be reminded of those traits. But did not Heidegger already glean this in his account of violence? For when we do violence to the everyday perspective, we are committing a violence that is directed inwardly; a violence directed towards ourselves as represented by that which is not ourselves, the they-self (das man), and a violence directed towards others, towards the they, which is represented and epitomized by that which we (through responsibility towards our own authentic self) can no longer resolutely accept about ourselves. This, for Heidegger, is a necessary violence needed so as to escape the inauthenticity of the they-self.

Perhaps, then, we ought appeal to Heidegger in order to help us see the full scope of what our violence and wrath towards others entails. I say ‘full scope’, because Levinas is partially correct, after all: we do wish to do violence, and even kill that which is altogether Other, since it is alterity itself, such as that elaborated by Levinas, which makes us almost as uncomfortable as we wantonly make ourselves! And it is Heidegger’s telling of this movement of violence (a violence directed at the self and turned towards others: das man, the they-self) that, in the presence of others, exists as an inward cruelty that becomes masochistic. As a result, one might be inclined to suggest that, conceptually, it is both masochism (Heidegger) and a kind of sadism (Levinas) that constitute the intentional grounds for how violence manifests itself and becomes concretized in our world. If we are to accept this, then the inadequacies of Levinas’ account of violence are ameliorated by the account of violence presented by Heidegger. As a result, thinking Heidegger alongside Levinas might offer us a richer and more developed picture of what motivates our violence.

---


2 I say ‘sadistic’, because Levinas’ notion amounts to a violence commissioned against that which lies outside the self, and is exemplified most profoundly, of course, in the anti-semitism responsible for the Holocaust.
The More Complete Picture Already Realized: Arendt on Violence

But did Arendt not already present us with this fuller exposition of what precipitates violence? In her *The Origins of Totalitarianism*, Arendt argues that race was the Boers' response to the “monstrosity” of Africa. The notion of race, according to Arendt, arose in response to seeing people who could not be recognized as like ourselves, not because of the color of their skin, but because of their relationship to the world they inhabited. They did not possess, says Arendt, anything like a constructed world of enduring objects, but continued to live as constitutive parts of the natural world. Behind racism, then, there is a nature/culture distinction (and, thus, a freedom/necessity distinction, as well).

This distinction, however, may suggest something far deeper. In her account, and following Joseph Conrad, there was something terrifying about these natural beings. That terror was likely due to the fact that they were *indifferent* to the claims of European civilization. Therefore, in Conrad’s view, from this European perspective, they were not *altogether inhuman*: thus, in slaughter and murder there was a form of self-affirmation. This can only be so because they—the “natives”—represent traces of nature in the human that have not yet been subjected to *world-making* (the construction of the human artifice). Or, rather, they represent traces of the human that are, in fact, always present. That is, what was terrifying about the natives is that they represented our own savagery, and, in exterminating them, we hoped to exterminate the otherness within ourselves. In other words, racist slaughter, the famous ‘exterminate the brutes,’ is the attempt to exterminate within ourselves anything that stood outside the existing comfortable categories of the notion of European world.

It might be said, then, that this European yearning is, in fact, a desperate attempt to exterminate the unconscious itself. This is certainly what seems to be the thrust of the Arendtian analysis. The natives are representative, in Arendt’s account, of that which is always unseen, always there, always on the verge of escaping, and, therefore, *intolerable*. So, we project onto them our own terror of our own worst instincts, and kill them, slaughter them, in order to satisfy ourselves that we are not *that*.

Furthermore, it is earlier in her work, that Arendt addresses what is, for her, the culpable naïveté of the Jewish people. If you live in an historical world, then you must, Arendt thinks, be a political creature. That is, living in history makes us all politically responsible. Therefore, an avoidance of politics is an avoidance of the brute fact of living in history, and to think that way is, for Arendt, a matter of culpable self-deception. This refusal of politics, though, has a pattern, and that pattern is the pattern of Jews making and keeping themselves a *separate* people (hence, the patterns of Jewish life itself). Given these patterns of Jewish life (of keeping themselves a separate people, of maintaining one’s Jewish identity, etc.), Arendt argues that the response of anti-semitism was not beyond the realm of normal human meanness or unimaginative resistance to the *foreign* and the *dissimilar*. That is, for Arendt, if you make an effort to keep yourself apart as a people, do not be surprised if ordinary people (with ordinary fears, worries and meannesses) despise you for
it (of thinking yourself different, better, apart or separate). You will be blamed by them, says Arendt, and they will hate you.

These two instances patently demonstrate why Arendt already understood that violence itself is precipitated by these two great moments of the human psyche (in spite of her well-known repudiation of psychoanalysis): that directed within (via her account of racism), and that directed without (via her account of anti-semitism). More importantly, for Arendt, none of that matters politically—this is, as we shall see, what truly sets Arendt apart as a thoroughly anti-liberal thinker.

The Modern Categories of ‘Liberalism’ and ‘Conscience’

But in view of this amalgam of ‘sadism’ and ‘masochism’ that schematically constitutes what we might see in ourselves as precipitous of violence towards others (what Arendt already sees), what precisely is it, representationally, that we intend our harm towards? In other words, when we direct our violence towards others, what is it that said others represent psychically? Would we be amiss in suggesting that, in terms of both ‘sadism’ and ‘masochism’, we wish to do violence to all that which does not accord harmony with the self? And if we grant this, is this self-harmony not, itself, a consequence of liberalism’s stake in the space of self-realization and individual fulfillment? ‘Should we all not have the right to actualize our own conception of the good life?’ asks the modern liberal. Is this not, then, the clearest expression, oddly enough (the irony therein will be considered in greater detail shortly), of conscience? For, Levinas’ claims in Time and the Other that the Other is disruptive of self-harmony, seem to unequivocally confirm that violence, is a thorough commitment on the part of our intentions to re-establish a sense of self-harmony for which we so desperately long.

Again once more, Levinas only elaborates one aspect of what engenders our violence, while Heidegger’s account offers us the element what is missing from Levinas’ account. Meanwhile, it is only Arendt who seems to offer us a thorough understanding of violence. Latent, however, within Levinas and Heidegger’s accounts of violence is an ineluctable dimension of conscience; an understanding of violence as a way one maintains a sense of self-harmony, self-realization, and tranquility via the mediation of conscience. For Levinas, ‘it is also not without importance to know that war does not become the institution of war in good conscience.’

Furthermore, for Heidegger, what he refers to as ‘wanting to have a conscience’ (Gewissenhabenwollen) is, in fact, the very basis of authenticity, as it is the resolute man who is guilty, knows he is guilty, and wants to have a

---


conscience through a reticence which will, indeed, allow the ‘call of conscience’ (*Gewissensruf*) to be heard.

Thus, as hinted to earlier, according Levinas and Heidegger, such a distinction assumes, oddly enough, and perhaps with great irony, that, at least in this sense, *both are unreservedly, all too liberal thinkers!* Having said that, however, we will shortly come to why, despite having most fully participated in this picture of violence, Arendt’s theory of conscience is, indeed, quite different from the others, as it offers far more than merely a fastidious return to a tolerable sense of self-realization, harmony, and conscience. In other words, it is anti-liberal.

The Relation Between the ‘Doer’ and the ‘Deed’: the Old Categories Revisited

But since our main concern lies in the establishment of non-totalitarian frameworks, let’s consider Levinas’ account of violence, the intent therein, and its place in modern society. For, what Levinas does not account for is the utter loss of the ‘thou shalt not kill!’ in the “enlightened” industrial age. This, after all, as Arendt, shows, no longer exists in the modern world, and this is, in fact, most clearly manifested in the death camps themselves. For Arendt, part of the threefold destruction of the human being (the individual, the moral, and the legal human being) that are actualized in the camps entails the destruction of the distinction between life and death, and the destruction of death as a meaningful end to a fulfilled life. As with Freud, for whom the 'thou shalt not kill' is the minimum taboo, the robbing of death as a meaningful event (as there are now, with the camps, worse things than death) destroys our capacity to act as moral agents. That is, it destroys the moral human being (therein lies Arendt’s notion about the development of utilitarianism as the moral philosophy of the camps). While Levinas relies heavily on the ‘thou shalt not kill’, he doesn’t account for the fact that it is precisely this commandment that, in our age, is *no longer available.*

In his account of violence, Levinas fails to account for the sheer disconnect between subjectivity and the external world, *between intention (conscience) and external action (in this case, violence).* That is, Levinas omits the fundamental disconnect specific to the modern world between ‘doer’ and ‘deed’. This would be, for Arendt, the direction in which Levinas *should* have gone when he argued that the ‘thou shalt not kill’ evinced in the face of the Other is not founded. For, as Arendt argues in *Eichmann in Jerusalem,* because morals, in keeping with the Aristotelian conception, are always constituted by *mores*—those social practices by which cultures establish moral categories, when social practices break down (because morals, in keeping with the Aristotelian conception, are always constituted by *mores*—those social practices by which cultures establish moral categories), moral categories lose their significance and authority; and all we are left with is leaving us with either the banal and routine administrative application of state doctrine—what might best be considered through Weber’s concept of ‘legal-rational authority’—which, when it alone is present, by itself can only provide us with his famous ‘iron cage’ of bureaucracy and rationality—*or the ability to judge.*
For Arendt, the failure to \textit{accomplish} the latter is an all too transparent indicator of a kind of Nietzschean frailty of spirit or mediocrity, as she often comments on the ‘strength to judge’—the strength exhibited, for her, on rare occasions by those such as Anton Schmidt.

For Arendt, it is precisely, the failure to judge, the ability to apply a kind of Kantian reflective judgment when moral concepts have lost their authority and weight, that most clearly elucidates where the ‘banality of evil’ is most often located. In other words, Levinas (and even Heidegger), in his thoroughly liberal understanding of violence, fails to account for the loss of the very element he viewed as having primordial significance in the possible preclusion of instances of totalitarian violence. Most significantly, \textit{that lost element is the constitutive connection between conscience and external action itself}. That is, not only is the ‘thou shalt not kill!’ wholly absent in a social order where moral categories have lost their authority, but the very conscience to which the ‘thou shalt not kill!’ is directed is itself disconnected from violence. Therefore, a deep and abiding concern with conscience, seeks refuge in the non-existent, while failing to appeal to what Arendt might rightly suggest is the only thing left standing in its wake: \textit{judgment}. As a result, Levinas offers us, not only an inadequate conception of what intends violence, but he fails to provide a properly sufficient non-totalitarian apparatus with which we can appeal.

Robert Bernasconi, however, suggests in his essay \textit{The Violence of the Face: Peace and Language in the Thought of Levinas} that Levinas was aware of this problematic. After all, Levinas says, ‘it is not without importance to know this so that war does not become the institution (\textit{institution}) of a war with a good conscience in the name of historical necessities.’\footnote{Levinas, E., (1996). ‘Peace and Proximity.’ \textit{Emmanuel Levinas: Basic Philosophical Writings}. ed. Peperzak, A. T., Critchley, S. & Bernasconi, R. Indiana: Indiana University Press. 170.} This, not only suggests that Levinas was in agreement with Arendt in condemning the notion of what Arendt calls ‘making history’, as it forgets the nature of human contingency and plurality, but, more importantly, in condemning the idea that abominable violence can often be waged with a clear conscience.

Levinas, according to Bernasconi, is keenly aware of the disconnect between conscience and external actions, especially those precipitating war. For Bernasconi, in some sense, Levinas already questions whether we can truly trust conscience. Referring to Derrida’s essay ‘Violence and Metaphysics’, Bernasconi employs a close reading of Levinas’ \textit{Otherwise than Being}, to suggest that Levinas did, in fact, appreciate the dubious nature of a ‘good conscience’.

But, as we have seen, Levinas is all too interested in a clear conscience. For, it is precisely through his conception of that towards which we intend our harm, that Levinas demonstrates that conscience itself is directly connected to what we desire doing violence to, thus giving us a liberal sense of self-harmonization. That is, even if the ‘thou shalt not kill’ (conscience) remains unfounded, it must be connected to external action for it to be heard at all. But in \textit{fully} recognizing that ‘God is dead’, Arendt discounts the connection of particular judgments to universal foundations—therein lies the necessity of
judgment. When considered in terms of according oneself a comforting sense of self-harmony and self-realization, conscience is, for Levinas, precisely that by which one is guided. In other words, what Bernasconi fails to see is that, though in some instances, it may appear that Levinas is, himself, suggesting a possible disconnect between conscience and external action, in articulating his concept of violence he employs the category of conscience itself to convey what precipitates violence! Since the liberal formulation of conscience as articulated by Levinas, is inadequate for our purposes I turn, then, to the question of how judgment in the specter of the breakdown between social practices and moral categories is best understood.

Arendt’s Repudiation of Conscience and, thus, Liberalism

With regard to social practices, Arendt can be seen as articulating a kind of Wittgensteinian theory of external action, --one which privileges the actual context with which one negotiates the world over intention. This lies in her equating the significance of an action with its inhabitation in the external ‘space of appearances’ through which a principle can be expressed rather than appealing to any simple correspondence between interiority (conscience) and exteriority (external action).

In her essay, Civil Disobedience, Arendt offers another rich description of conscience by differentiating between ‘civil disobedience’ and ‘conscientious objection’. Using Socrates as a model, Arendt claims that conscience is not political, because conscience functions solely in accordance with my own self-relation. In other words, when I act conscientiously, I do so, in order to foster an agreement with myself. As understood by Arendt, conscience, operates negatively; it tells me what I must not do to avoid self-destruction. For Arendt, if I did not abide my conscience commanding me not to rape or kill, I would have to live the rest of my days with a rapist or murderer: viz., myself; and this just is self-destruction. Therefore, we listen to the voice of conscience, because if we were to act against our ‘ego ideal’, we would be forced to live with that contradictory, villainous creature for the rest of our days. Who, in “good conscience,” would wish that upon themselves?

The power of conscience is the power of striving to bring myself in accord with myself, so as to be friendly and accommodating to myself. For Arendt, however, what brings one into harmony with oneself can be any content whatsoever. That is, morality and conscience can be occupied by any content. Conscience, therefore, is simply the subjective realm of self-harmonization, and conscience places the benevolent cultivation of the self, one’s moral health and prosperity, above that of the state. Conversely, Arendt’s favorite Machiavellian statement (which serves, against conscience, as a kind of archetype for political action), ‘I love my native city more than my own soul,’ operates as a way to conceive of external action which doesn’t lend primacy to self-harmony and self-realization. That is, it just like Machiavelli’s statement, Arendt’s theory of conscience is anti-liberal.

---

As what is of relevance politically is what, within the ‘space of appearances’, can become just one more opinion, conscience is as moot politically as the philosopher’s truths. That is, for something to be accorded political weight, it must be removed from the realm of moral certitude, and enter into the ‘space of appearances’. ‘Thou shalt not kill’ is, on Arendt's reading, little more than a philosopher's truth, a “moral certitude”. WSo, while not suggesting that one disregard their conscience, Arendt is claiming that, since it is not premised on plurality, abiding by one’s conscience alone or acting as a ‘conscientious objector’, in and of itself, is not a way one can be political. Moreover, conscience alone is insufficient in a plural world, since there is an interdependent obligation that supersedes it, and takes precedence politically. Conscience can enter into ‘the political’ once it’s prepared to become appearance and opinion whose quality is to be judged by the strength, not of conscience, but of the collectivity adhering to it—in short, once it is prepared to be equated to power.

Thus, a liberal framework, of the sort that Levinas employs against the horrors of totalitarian violence is one that relies on the connection between conscience and action, a connection between doer and deed. A more thorough understanding of the modern world must recognize the disconnect between these which results from the loss of moral certitude. Thus, the only possible response is an anti-liberal framework, which, of necessity, cannot rely on conscience. With this, then, we turn to the Arendtian model.

‘The Political’: A Fundamentally New Relation Between the ‘Doer’ and the ‘Deed’

For Arendt, it is ‘the political’ which is not premised on a uniquely modern disconnect between the doer and the deed that provides an adequate bulwark against totalitarian collapse. The political stresses the appearance of individuals in a shared space where, through speeches and deeds, plurality, uniqueness, and human excellence are allowed to flourish. Inside the domain of the political, one reveals oneself, meaning, and the world. The political functions as the transcendental ground for meaningfulness, and reveals, according to Arendt, that there are some things which are more important than merely staying alive—a notion which crystallizes in the death camps via its utilitarian morality which does not—indeed, cannot-- recognize the modern disconnect between “subject” and action. Through this shared ‘web of appearances’, we find that not everything is possible: we cannot ‘make history’, since ‘making history’ contradicts the contingency inherent in the universe—a contingency that we attempt to deny, when we suppose that we can be sovereign over the meaning and responses to our actions. The political is the domain that appreciates the novelty and new beginning that each individual person represents in their natality; a uniqueness that can only be recognized by others when it is revealed as opinion through storytelling, and then collectively judged in a structure appropriate to it. Arendt’s call is a demand for those structures, and for the plurality which can become manifest within them. In Arendt's terms, Levinas's domain of morality, does not privilege the fact that ‘men, not
Man, live on the earth and inhabit the world.\(^\text{1}\) Rather, it privileges one’s own subjectivity as a moral agent alone. By denying plurality and, thus, freedom, through an appeal to conscience, we deny appearance and opinion with others in favor of hidden subjectivity and sovereignty. According to Arendt, this thoroughly modern impulse has resulted in the disastrous events of the 20\(^{th}\) century.

Thus, as Arendt shows us, it is, counter-intuitively, liberalism’s quintessential element, conscience, that gets lost in the modern world. For it is precisely conscience that can be occupied by any content whatsoever. As with Eichmann, whose conscience was troubled by the “immorality” of Himmler who wanted to broker a deal whereby the Nazis would cease the murdering of Jews, conscience can compel us to commit murder just as easily as it can compel us not to commit it. How, then, are we to construct a non-totalitarian schema which might preclude violence? A liberal culture defeats the attempt by denying itself the connection between conscience and external action without which it is impossible to preclude violence in even the most extreme cases such as the Holocaust. Therefore, in a liberal culture paradoxically bereft of perhaps its most exemplary characteristic, the connection between conscience and external action, if one is to attempt to construct a non-totalitarian schema with which violence might be precluded, they must not attempt to employ that very framework of which it is now devoid! This, however, is precisely what Levinas’ conceptualization of violence, in the end, accomplishes by failing to appeal to, perhaps the only alternative to the routine administrative application of state doctrine left in the modern world: the ability to judge. It is in this sense, that Arendt, not only prefigures the apolitical perils of self-realization and self-harmonization that conscience, accords, but a more thorough and well-rounded picture of violence, as well. For, it is this very picture that reveals the perils in Levinias' and Heidegger's accounts of violence. Arendt, goes beyond this limitation, because she instructs us on the helplessness of conscience itself. In this sense, Arendt is the only one of these three great 20\(^{th}\) century thinkers who is thoroughly anti-liberal.

**Selected Bibliography**


Despite the somewhat notorious acrimony in their mutual regard for one another\(^1\), and the alleged polarity respective of their views, Wittgenstein and Gödel share a number of properties, not the least of which is the fact that the philosophies of mathematics propounded by both are consistently disparaged by contemporary philosophers of mathematics. Wittgenstein's apparent indifference toward contradictions and his insistence that proof is the extent of meaning continually elicit resistance if not disregard, while Gödel's assertions of extra-sensate truth are characterized as a sort of embarrassingly naïve Platonism and dismissed. In the following, I seek to show that these two alleged theoretic poles are in fact approaching a single philosophy of mathematics, and to describe an overlooked and meaningful interpretation of mathematical truth which lies just at that point.

**Merely Proof**

Wittgenstein sought to demonstrate that proofs are not revelatory in manner, but rather extensions of a stipulated mathematical rule set, and thereby constructive and self-justifying. Mathematical truth--whether theorems of arithmetic or type theory--thus becomes a matter of sheer invention rather than revelation. Mathematicians accordingly engage in a an elaborate game of sorts,

\(^1\)For an insightful investigation of their mutual regard and of the personage of Gödel in general, see Goldstein (2005).
replete with rules, any one of which is defined by its perspicuous derivation from prior stipulation or derivation.

If we take this to be the case, it clearly follows that the proof supplies meaning for its conclusion. Pressing an invented mathematical proposition for a 'meaning' external to its construction is admittedly delusional--analogous to asking what it means that bishops can only move diagonally on a chessboard or to seeking a principle to explain why there are only two white knights.

What is somewhat less easy to accept is the notion that mathematical truth is of this nature in the first place, and Wittgenstein devotes considerable energies to allaying the philosophical disquiet otherwise entailed by this assertion--and rightly so. Mathematicians certainly speak of 'finding' a proof of Goldbach's Conjecture\(^1\), for example, not of 'inventing' one. Even philosophers of mathematics--who often for better or worse pay no mind to the words chosen by mathematicians--argue about whether mathematical truths are discoveries of the implications of logic, the implications of our own formalities, or implications of our prior mental constructions.

Wittgenstein disagrees (granted, all three of these more popular interpretations--Logicism, Formalism, Intuitionism--can reasonably be asserted to have been influenced by Wittgenstein, but each breaks from him at some point, and he from them). He posits and advances the notion that the so-called discoveries of mathematics are in fact no more than the continual rule-building in a motley of mathematical games. For indicative examples he favors deceptively simple proofs of arithmetic statements, often in pictorial form such as the following, to show that even a proof of the statement '2 + 2 = 4' is merely a matter of stipulation (Wittgenstein (1983, §15). He reacts to the supposition that two groupings of two units (see Figure 1) is a basic visual proof of the independently-existing truth that 2 + 2 = 4 with the analogous statement that three overlapping groupings (see Figure 2) is a relevantly similar proof that 2 + 2 + 2 = 4. That is, the former demonstrates that two groupings of two units come to a total of four units, whereas the latter demonstrates that three groupings of two come to a total of four units, and that the apparent contradiction lies in a relatively arbitrary methodological decision. Thus there is a sense that both figures make equally valid claims to demonstrating simple arithmetical truths.

Mathematicians and philosophers alike react negatively, needless to say, to the declaration that Figure 2 is anything of the kind. I admit that indeed my own first encounter with this example led me to believe that Wittgenstein was practicing the worst kind of sleight of hand masquerading as philosophy, but let us consider carefully just what is so abhorrent herein.

Can we imagine a world in which Figure 2 and its implicated arithmetic are more sensible than the arithmetic following from Figure 1? I believe we can. It would be rather like a world of overlapping pieces, in which it made sense to count from above. If we were surrounded by such a world, one can imagine where two two-pieces would always yield three X's (because they

\(^1\)Notably, this unsolved problem (proving that every even integer greater than two can be written as the sum of two primes), posited by Christian Goldbach in 1742 is of recurring philosophical interest to both Gödel and Wittgenstein.
needed to be connected, say, on an X) and two 3-pieces would always yield 5 X's. We would undoubtedly speak of the operation as 'addition', and the sum of two and two would be three, the sum of three and two four, etc.

We can imagine an entire system of mathematics being built around this operation (presumably something akin to the set of natural numbers would follow from 'adding' two to each preceding term rather than one). We may even be convinced in this world that we were discovering the mathematical implications of 'addition'. Furthermore we can imagine that were it the case that humans inhabited such a world from time immemorial, we could even go so far as to construct a 'false' addition where the sum of two and two (through manipulating hypothetical structures) was four--a system that would no doubt evoke dismay in the world in question.

The reaction to this is, reasonably, that we are not in a world of overlapping pieces connected by X's, that in this universe 2 + 2 'actually' equals 4. Actually? Could we not assign the overlapping-pieces conception to arithmetic this very moment and start doing calculations with it? It seems we could, there is certainly nothing stopping us. The only relevant difference would be that the new arithmetic would (likely) turn out to be less useful in our world than the one we have heretofore assigned. Note, it is not that the new would be impractical whereas the old is practical, it would simply be less so.

So the best argument for the present system becomes that it is useful, that all of our rules, derivations and proofs are usable in practical reality, more than their competitors, and perhaps that our reality is such that these systems are more useful than others. If this is the case, then our mathematics must truly be an ongoing construction project--or many such projects. Regardless of our inclinations toward discovery, we are restricted to invention, and proofs and derivations are the self-imposed limitations of mathematical meaning. It is for this reason that Wittgenstein advocates the senselessness of inquiring as to the extant truth of something that is yet beyond proof, such as the continuum hypothesis.\(^1\) If a mathematical proposition has not been derived or negated by the extension of our mathematical rule set, it is not something about which we can meaningfully inquire.

Thus a proof 'shows' nothing, but instead effectively supplies meaning to itself, and is tested through usefulness. As to whether there might be some ultimate usefulness, so to speak, in inventing just the right collection of proofs (and hence such invention be considered a discovery after all), it is a question that will bear repeating after examining further the notion of proof and ultimacy.

It is easy to see why Gödel's incompleteness theorem is cast as something quintessentially un-Wittgensteinian, and vice versa. On the face of it, Gödel seems to be saying that proofs aren't everything, while Wittgenstein just assured us that they were. In fact, the super-formalism attributed to Wittgenstein and the super-realism attributed to Gödel seem to be opposite ends of a long and variegated spectrum of philosophies of mathematics. To

---

\(^1\)The hypothesis, advanced by Georg Cantor, that there is no set whose size is strictly between that of the integers and that of the real numbers, has not been proven or disproven.
further set these discussions of proof apart, each Philosopher, for his part, calculatedly eschews the other: Wittgenstein famously sets out 'not to refute Gödel, but to pass him by' (Wittgenstein 1983, §71), while Gödel characterizes Wittgenstein's work as 'nonsense'. It should be noted, though, that Wittgenstein spends much energy focused on that which he wishes to 'pass by', and that Gödel accuses Wittgenstein of pretending not to understand his incompleteness theorem—a strange accusation indeed toward somebody one considers senseless.

In order to demonstrate the possibility of a latent similarity in these two attitudes toward proof, it is helpful to investigate what each means for systems of provability. Gödel shows us unequivocally that any given system of provability is fated to incompleteness by its own nature (Gödel 1992). The incompleteness proof demonstrates that we can even invent external systems or rules just to patch these gaps, but that all of our efforts will only lead to a collection of rules and systems which are each unable to reach beyond themselves, or even fully extend to themselves. In essence, mathematical logic is doomed to be a motley of provability systems, none of which can ever make a claim of ultimacy, and we should keep in mind the limits of such lest we overextend our reach with regards to what is provable. Gödel insists that provability is just provability, and Wittgenstein agrees.

**These Metamathematics Cannot Be**

'There isn't any metamathematics,' writes Wittgenstein (1978, §290). Of course not. Given that he has shown mathematics to be a set of rules one builds upon a sort of enterprise, any theory attempting to explain the nature of this process is doomed to as much meaninglessness as meta-chess theories might have. The very notion of metamathematics becomes accordingly incoherent.

Of course, the notion that mathematics consists in the invention of rules and constructs is itself a statement about the nature of mathematical truth (albeit perhaps a negative one—a statement concerning what mathematical truth isn't), and so perhaps Wittgenstein would have been better served by saying, 'There is no metamathematics other than the fact that there is no metamathematics,' or something to that effect. Engaging in the attempt to prove positive statements regarding the nature of mathematical truth is, according to Wittgenstein, doomed to failure.

Attempting to prove positive statements regarding the nature of mathematical truth is, according to Gödel, doomed to the very same failure, or a strikingly similar one at the very least. There are limits to provability, and we cannot devise a system in which what is provable actually corresponds to what is true. In describing the nature of any given system, we are forced to invoke (invent?) another system, and in this regress questions of truth will inevitably outpace provable consistency. Metamathematical language is thereby a sort of self-deception, as so-called metamathematical statements will only become provable as they also become the object of description rather than the source. Truly metamathematical statements are unreachable.
Of course, Gödel's limitation is a discussion of the nature of mathematical truth (albeit a negative one). So perhaps I should revise accordingly: Truly metamathematical statements other than this one are unreachable.

The Usefulness of Contradiction, and the Meaning in Usefulness:

So it seems momentarily that there is no ultimate usefulness, at least not an attainable one, and so we are left with empty, arbitrary formalisms. We can only construct meaningless and fragmented systems, which we cannot even adequately describe without constructing other meaningless and fragmented systems. Altogether a despairing state of affairs, granted, but this is not the point at which either theory stops.

Although we have dispelled the notion of ultimacy in our systems, neither explanation has at all advocated that all systems are equivalently inadequate—we have thus far only been guaranteed that no system is without inadequacy.

Let's return to the world of overlapping pieces and the ensuing arithmetic. Imagine a delusional but gifted mathematician perceives our universe as such, and starts creating the applicable mathematical systems this very day. In time he has worked out properties and proofs for all manners of derivation in this system where $2 + 2 = 3$. We admit that perhaps we are not willing to say he is wrong in that statement—that is, the statement is as true in his system as $2 + 2 = 4$ is in the one with which we are more familiar. Furthermore, upon consideration it seems that his system couldn't be said to be false, at least not in some way by which we regarded ours as true; especially since Gödel assures us that the articulation of a statement demonstrating the truth of an entire system is fated to be a Sisyphean task.

So what do we tell this deranged mathematician? We tell him that we have a different system, and that our system is more useful. We compare what can be accomplished practically with each, and presumably emerge victorious. This seems to be (and is) a statement about the respective systems: that by the admittedly humble standard of usefulness, one system (ours) outstrips the other (his); but it is more than just that, it is also a statement about the nature of our mathematical reality: that it is the sort where a mathematical system based on statements like $[2 + 2 = 4]$ serves us more adequately than a system based on statements like $[2 + 2 = 3]$. For we are driven to ask why one system is superior to another with regards to usefulness, and in doing so we are actually seeking one of the statements that Gödel and Wittgenstein just prohibited. The only allowable answer is that one is more useful because the nature of reality—which we cannot articulate—is such that our constructed mathematical system is a fairly useful one.

A circularity, certainly, but not a hollow one. In fact, the circularity here is expected, since we necessarily can answer that question no other way. We could rephrase the answer thusly: 'We cannot say it, but there is an answer to that question.' The fact that an answer exists—even an inarticulable one—is extremely important. We cannot ask the question expecting a proposition-by-our-rules in return, for we were just guaranteed that all such systems of rules are intrinsically inadequate, but we do receive the meta-answer that there is an answer. Effectively the very fact that one theory is more useful than another is an existence-proof for some quality of mathematical reality. Which leaves us
with one hesitation: did we just succeed in stating a truth about the nature of mathematical truth? Fortunately, the answer is actually in the negative.

It is tempting to at this point say, 'something is true and our system (say, ZFC\(^1\)) reflects that.' We could then refine this statement about the the reality of mathematical truth into a rather proper one along the lines of 'it is true that there are mathematical truths by nature, and they are more like the truths of ZFC than they are like many other systems.' However, this would be mistaken, as we are already playing a language-game when we get as far as 'something is true,' as Wittgenstein assures us, and the language-game we play with truth-value bears no special exception to the limitations that have been prescribed to all such systems. This may seem to be an absurdly serious interpretation of these limitations, but consider the varying systems of logic and their applicabilities: bivalent logic certainly has claims to usefulness, but so do more recent reworkings like paraconsistency and dialetheism. As soon as we predicate the nature of reality by means of invented language-games, we are overextending the reach of our systems of provability.

Hence we cannot begin by saying 'something is true.' 'True' belongs to one or several of our invented systems, fit badly onto that which makes one system more useful than another. We must instead begin by saying, 'something is something'. I realize that is seems as though we are met with utter absurdity here, but it is not so. In effect, philosophers of mathematics have been wondering whether something was indeed something since Pythagoras and probably before, and we can answer that query by finishing our thought. Something is something and ZFC is fairly useful by virtue of whatever that is. Something akin to a mathematical proposition is something akin to true about the nature of our reality, and although we cannot articulate these para-truths, we can contrive (and have contrived) systems which correlate to them with decreasing, but never nonzero, inadequacy.

Now we see why Wittgenstein was famously un-averse to contradictions\(^2\). He admitted they were one thing to be avoided, but also urged us not to over-avoid them, as he insisted we've been doing all along. If our systems are doomed to inadequacy, it should come as no surprise when contradictions crop up here and there. If we can reinvent such that usefulness is retained but the contradiction is stripped, so much the better, but in some cases contradictions may merely be a symptom of the bad fit between our invented systems and the aforementioned para-truths they pretend to represent. In fact, since we are assured of this bad fit from the start, we should expect inconsistencies to appear and may even be able to use them to our advantage (the paraconsistency logician is already familiar with this idea, as was Gödel, presumably, since his Incompleteness theorem is rooted in paradox).

This is exemplified by the Axiom of Choice and the Banach-Tarski paradox\(^3\). Traditionally, we have asked whether the axiom of choice can be

\(^1\)Zermelo–Fraenkel set theory with the axiom of choice (commonly abbreviated ZFC) is the most common form of axiomatic set theory.

\(^2\)Among many other similar expressions can be found, 'There is one mistake to avoid: one thinks that a contradiction must be senseless' (Wittgenstein 1989, §V – 14).

\(^3\)Asserted by Stephan Banach and Alfred Tarski in 1924, the Banach-Tarski paradox or Hausdorff-Banach-Tarski paradox states that by using the axiom of choice it is possible to take
taken as true or not. The problem, of course, is that if we try to go on without it no end of troubles appear, and if we take it as true, we apparently end up with an unresolvable paradox. The interpretation here presented directs us not to ask whether it is 'true' (i.e., 'true of reality'), but instead to ask whether it is useful, and whether we can reinvent something as useful but less paradoxical—admitting of the possibility that there is not a satisfactory resolution to that inquiry. Paradoxes provide serious foundational difficulties for a conventional realist (or even an anti-realist), but after resigning ourselves to the intrinsic inadequacy of our invented systems to accurate reflect whatever true-like proposition-like things are driving the relative usefulnesses of our systems, paradoxes become merely a part of the landscape. Encountering a contradiction in the implications of any given invented system will be expected and may even prove useful in the development of the system, the development of other systems, or in understanding the way that various mathematical inventions approximate those qualities of nature which drive one 'manner of notation' to be more useful than another.

Φυσις Κρυπτεσθαι Φιλε

Whether the 'truths' of mathematics are discovered or invented has been a difficult question for centuries. On the one hand we seem to forsake objectivity if we choose invention, but on the other we seem to make overly bold epistemological claims if we endorse discovery. In effect, the interpretation of mathematical truth here sought acknowledges the existence of objective qualities of reality which dictate the usefulness of our contrivances but disallows us from articulating any truly descriptive statements regarding the nature of these qualities. We invent, then, but our inventions are constrained by instrumentality which is determined by an objective state of affairs. We can and should seek to improve our systems with regards to usefulness (and hence with regards to their correlation to these indescribable qualities of reality), but since nonperfect correlation is guaranteed we cannot but expect some degree of inconsistency, incompleteness, and paradox. Furthermore, we cannot expect to say more of the objective mathematical qualities of reality than that our reality is such that a particular invented system, albeit inherently limited, is useful in describing it.

---

1An aphorism of Heraclitus, usually translated as 'Nature loves to conceal itself' or 'Things keep their secrets'.

---

a solid sphere in 3-dimensional space, cut it up into finitely many (non-measurable) pieces and using only rotations and translations, reassemble the pieces into two spheres of the same radius as the original.
References

Part IV

Metaphysics, Epistemology and Philosophy of Language
Different labels are used in discussions concerning realism. I will use some of them, but only according to the definitions given. This applies particularly to the expression ‘Internal Realism’. I aim at systematic argumentation and will not give an interpretation of Putnam’s views.

Different varieties of realism can be distinguished according to axioms of realism which they support:

(R1) Reality exists independently of our cognitive faculties.
(R2) Independent reality divides into entities, which have structures and stand in relations to each other.
(R3) To some extent we have epistemic access to reality as well as its structures and relations.
(R4) (R3) plus the claim that our access to reality is limited by our cognitive faculties (particularly our linguistic and perceptual faculties).
(R5) Notwithstanding our modes of cognitive access, the structures and relations of reality are represented in our cognitions (particularly in our representations).

(R1) expresses ontological realism regarding the existence of reality; a position that very few have ever seriously doubted. (R2) expresses ontological realism regarding the structures and relations which reality exhibits. (R2) does not entail that we know of these structures and relations. One can ascribe (R2) to a philosopher if she speaks of a plurality of unrecognizable ‘things in themselves’. (R2) claims that we are not the makers of reality and its structures.

(R3) is stronger than (R2) in professing not only the ontological independence of reality but also epistemic access to these structures. (R3) I call ‘strong realism’. According to (R3), this epistemic access is given ‘to some extent’. The vagueness of this phrase is inseparable from the realist’s
epistemological views. The relations of our representations to reality pose the problem of the epistemic gap: our representations shall concern reality but we cannot guarantee this. Therefore we cannot say how much of reality we recognize. Just in case realism is correct, we cannot give an ultimate proof of it, because we can only reason from our side of the epistemic gap. Therefore a realist must be content with epistemic access ‘to some extent’. Often (R3) is simply called ‘realism’. But (R3) is, in my view, only the common ground of two more advanced realistic positions.

I call the first of these, (R4), ‘internal realism’. Thus, on my view, internal realism is a species of strong realism. The second, (R5), is what I call ‘external realism’.\(^1\) [(R5) is often referred to as ‘metaphysical realism’. But this name is ambiguous and leads to confusion. For example, (R3) is sometimes referred to as metaphysical realism. If this is so, then (R4), my internal realism would count as a species of ‘metaphysical realism’, a view I reject. To avoid this potential problem, I don’t use the phrase ‘metaphysical realism’.]

While few have ever doubted (R1), it is difficult to find philosophers who subscribe to (R5 with no reservations. Any view which has a commitment to a distinction between essential and accidental \(de \text{ } re\) qualities of entities might fit under this heading. Similar remarks apply to ‘modal realism’. (R5) could even be expressed by a picture theory of truth or meaning. (R5 allows for the fact that a representation fits reality and we have no idea about this.

Internal realism keeps the gap between our opinions and reality, further, it implies a certain sceptical possibility of systematically distorted cognitions. Even if we assume, with (R3), that this possibility doesn't obtain, there is no proof of this. That we cannot step outside of our cognitive formats and access reality undisguised is the internalistic element of internal realism. The sceptical possibility is the other side of the ontological realist’s coin. To exclude scepticism threatens to diminish reality to human measure. This over-estimation of our epistemic abilities also threatens the external realist (R45 if he is epistemologically very optimistic (e.g., proposing a picture theory). A quarrel whether internal realism (R4a is, therefore, more realistic than external realism seems scarcely productive. It seems more adequate to describe internal realism as less realistic inasmuch as it makes truth an epistemological concept.

Internal Realism, nevertheless, is a version of strong realism because of the following.

(Thesis 1) Epistemological positions which deny (R3) fail, since they cannot justify any theory of meaning in which the definiteness of the meanings of our linguistic expressions is maintained.

*Justification:* We make assertions. We use these assertions to describe what is the case. Statements (or sentences) which are claimed to be true are statements of a specific language and use the vocabulary of this language. They

\(^1\) (R5) is often referred to as ‘metaphysical realism’. But this name is ambiguous and leads to confusion. For example, (R3) is sometimes referred to as metaphysical realism. If this is so, then (R4), my internal realism would count as a species of ‘metaphysical realism’, a view I reject. To avoid this potential problem, I don't use the phrase 'metaphysical realism'.

\[182\]
describe, by means of the expressions occurring in them, what would be the case if they were true. Statements which use different vocabulary describe, except in some cases of synonymy, different facts. The specific vocabulary, that is to say, the meanings of the words used, makes all the difference and determines which statements we consider true and which not. To illustrate this, consider the following.

Assume the statement ‘F(a)’ is true. If we can intersubjectively refer to a, being describable as ‘F( )’, we justify or verify [(in a weak sense of ‘verification’)] the claim that F(a) is the case. We thereby claim that the thing which we consider to be F possesses the features (the structure mentioned in (R2)), to which the meaning of the expression ‘F( )’ refers. The realistic interpretation of this procedure is this. Our claims to knowledge of a reality beyond the confines of our representations are expressed here. The definitions or conventions by means of which we refer are believed to correspond to the composition of the actual entities. We refer to reality by means of language and try to reach an agreement about what is objectively the case.

This procedure of reaching agreement and the collective and individual practices based upon it are more or less successful. The object a might not be exhausted in its features by describing it as ‘F( )’. But we assume with the truth of ‘F(a)’ that it has, at least, this feature in reality. Definiteness of meaning is founded on this correspondence. Definiteness of meaning presupposes that different expressions have conditions of application separated by their meanings, and that it is, for example, clearly distinguishable whether to use the expression ‘round’ or to use expression ‘square’. The condition of this possibility is (R3) (strong realism). (R2) as a component of (R3) explains the first aspect of definiteness of meaning as follows: descriptions of facts (sentences) are definite because the expressions composing the description refer to components (parts) of reality. The difference of the reference situations guarantees the difference of the meanings of the descriptions, since an aspect of these meanings is reference. Linguistic expressions are referentially or extensionally definite according to (R2). Assuming referential definiteness we have to endorse (R2). We still have to say something, however, about our ability to use expressions in the appropriate situations. The definiteness of use and application refers us on to (R3). In this respect the intension of an expression (the meaning in the narrower sense) consists in instructions for the application of the expression mentioning some decisive features or criteria to be fulfilled. We must in some way or other be able to decide or discriminate the fulfilment or non-fulfilment of these criteria to use the expression definitely. This ability might be instantiated in explicit linguistic reflection or perceptual or sub-doxastic processing (that is, it might be a not directly conscious process of information processing). We must, however, be able to recognize the structures and relations of reality to some extent, whatever way this happens.

This is (R3). We intentionally employ specific expressions in distinction to others. That the use of a specific expression makes a difference in what is said and that it is founded on real differences in the situations of application we cannot deny without dissolving our intentional linguistic behaviour in an arbitrary utterance of some expression or other. That is: (R3) must be true if we
are able to speak a definite language. For (R3) we can give another short argument:

1. Making statements works (sufficiently well). (undeniable fact)\(^1\)
2. Intentionally stating something implies making distinctions. (by definition)
3. Non-definiteness of meaning implies inability to make sufficient distinctions. (by definition)
4. Definiteness of meaning. (from (1)-(3))\(^2\)
5. Intentional application of an expression is successful (if and) only if we employ the expression only on a specific occasion to which we have cognitive access. (Meaning Principle)\(^3\)
6. We have cognitive access to the situations we make statements about.

The consequence, (6), again expresses strong realism. The argument establishes some access to reality, but this is enough for (R3) to be true. Given some access and lacking justified doubt we may claim access as an ordinary phenomenon.

Two supplements to these two arguments supporting (Thesis 1).

(I) If our expressions refer to structures of reality, we shall be able to develop an ontology of facts (states of affairs), facts being the building stones of reality (the physical correlates of true statements/sentences). Such a robust notion of fact supplies the realist with a relatum for the relation of correspondence.

(II) Putnam’s Theorem (Putnam 1981, Appendix): Putnam’s main argument against a reliable assignment of expressions to their extensions and for his internal realism. Putnam employs the Löwenheim/Skolem-Theorem stating that concerning any First Order Language an assignment of expressions to referents can be permutated giving us non-intended models. This application of Löwenheim/Skolem to natural languages should be rejected:

(i) Probably no natural language is a First Order Language.
(ii) Even if natural language is a First Order Language, the assignment of meaning in a natural language will not be adequately modelled by an interpretation function (Bremer 2005, pp. 241-42).

Thus, Putnam’s Theorem should not hinder us from being strong realists.

---

\(^1\) If you try to deny this, you immediately refute yourself.
\(^2\) Or, for the intuitionist, Non-Non-Definiteness of meaning.
\(^3\) This principle, of course, has to be argued for. I tried to do so in Bremer 2005. Note that this principle and the intuitionistic reasoning employed in the argument should be acceptable to so called ‘anti-realists’ in the style of Dummett or Tennant.
We Cannot be External Realists

A pure coherence theory of truth, in which sentences are justified only with respect to their mutual coherence, violates the realism I have just defended like a ‘radical’ constructivist position does. A pure correspondence theory, however, makes truth something beyond our cognitive faculties. The epistemological advantage of internal realism consists in avoiding both disadvantages, but external realism (R5) has to be given up.

If truth were a completely non-epistemic concept there could be entities and qualities which we talk about without us ever being able reasonably to state this correspondence of language and reality. Truth could not even be prima facie established, since there would be no way of introducing any criteria of truth: to introduce such criteria we would have to judge them in their reliability against other candidates, but if we were not able to access truth in the first place, we could not establish anything as being truth conductive. Nevertheless, we keep on claiming things to be true. And to claim that something is true requires justification. If a speaker A asserts that p, she claims that p is the case, that it objectively obtains and is not merely A´s opinion. To claim objectivity concerning p makes not only the difference between mere belief that p and p being the case, but also claims that p will be the case for any speaker B of the linguistic community. This is exactly what ‘objective’ means in contradistinction to ‘(merely) subjective’. Two questions then have to be answered.

1. How is it possible to distinguish mere opinions from opinions to which facts correspond?
2. How can one decide between A´s assertion that p and B´s assertion that non-p?

There must be means to answer these questions if understanding is at all possible. For any attempt of communication about what facts obtain to succeed, these questions must have been answered already. The means sought for are reasons. We accept, ideally, those assertions which have been better justified than their competitors. To give reasons is beneficial to establishing the truth of a statement which claims that a certain fact obtains. We are interested in assertions because we establish by their means that which we shall consider as being the case. To doubt that we are seeking objectivity is a move in the language game that undercuts itself. We call statements ‘true’ or ‘well-founded’ to distinguish between mere opinions and facts. We are thus lead to (Thesis 2) Truth is conceptually tied to justification.

This concept of truth aims at correspondence but connects this idea with criteria of consenting to statements, and reasons. We have, therefore, reached a dual aspect theory of truth which corresponds to internal realism.

---

1This justification is no final verification in the sense of some absolute external access to truth, which no realist would allow for. What is meant by ‘truth conductive justification’ and ‘verification’ in such a narrow sense is to engage in putting forward reasons why some belief is true, and which so long as being undefeated allow for assuming the corresponding facts to obtain.
The above argument presupposes the following.

1. That we take seriously the claim that statements are/can be true (i.e., we should not reinterpret it as actually being assertability, constructibility or some of the kind).
2. That it must serve a purpose, when we claim truth, and that the purpose of asserting something to be true is truth itself (as an epistemological basic value) or something which we need it for (e.g., successful manipulation of our environment).
3. That we can realise this purpose; otherwise we would not persist in trying.
4. That we know about the contrast (merely) subjective vs. objective.

The acceptance of (d) hinges on conceptual analysis being possible in a minimal extent at least. Accepting (a) expresses a conservative attitude. Whoever wants to substitute assertability for truth can do so, but then he has already incorporated an epistemological feature into the concept of truth. My argument adresses realists who have not yet done this. The acceptance of (b) and (c) depends on taking seriously evolutionary or transcendental functionalism searching for the conditions of possibility of something which has worked historically. I take (a) to (d) to be acceptable and rather weak assumptions, indeed.

The above argument centers on the normativity of meaning: since we should use an expression only on some occasions, we have to justify that our using the expression concerns an occasion of the appropriate type.

A shorter version of an argument for (Thesis 2) might concern the felicity conditions

1. I assert p if and only if I claim p to be true.
2. For any speaker, sentence, event: if event e consists in A claiming p to be true and e is successful, then it is possible to verify/justify p.
   (Felicity Condition 1)  

After some steps we arrive at C. For any sentence p: if it is not possible to verify/justify p, then there is no event e such that there is a speaker A such that e consists in A’s claiming p to be true.

This conclusion makes epistemically transcendent statements/sentences unassertible. And if external realism depends on sentences which are unassertible we should give it up.

\[\text{1This condition would, of course, have to be argued for. A theory of speech acts might do so. Asserting would lose its point if I am not taking responsibility for what I assert. Without the requirement of justification and justification being decidable I could assert just anything.}\]
The Concept of Truth that Internal Realists have to live with

The last paragraph set up a dual aspect theory of truth. There is a connection between reference and procedures of our cognitive faculties. Reference takes some of our forms of representation as starting point. Therefore truth - concerning statements referring to facts - entails an aspect of justification, whatever epistemology of justification one prefers. Our ways of speaking, though, do not ‘make’ reality. Epistemic procedures which are conducive to success mustn't run completely against the data. Rather we adapt our languages and methods. The ‘idealistic/constructivist/Putnam-Goodman-style’ inference from the particularity of different modes of representation to the dependence of the concerned structures and entities on consciousness is simply wrong.

The internal realistic concept of truth can be stated thus:

\[(TR^4) \text{ A statement is true if and only if it is an intersubjectively justified agreement that it has to be inserted in our best-founded frame of reference, and this frame of reference corresponds to reality.}\]

An argument (a reason put forth) tries to establish that some statement has to be included into our best frame of reference. It forces itself upon us given our other commitments. Therefore somebody made the respective assertion. That some statement might be true (could be coherent with our best frame of reference) is not enough as it does not force that statement upon us.

To say of a statement that it is true is to ascribe this material quality defined by \((TR^4)\). Only statements of languages in which our best-founded frame of reference can be formulated can have this property. ‘Best-founded frame of reference’ is meant to denote a frame/theory that could be formulated in a given language even if we have not yet done so. We might improve our modes of reference, but the best-founded frame of reference is determined whether we know it or not. And whether some statement is part of this frame of reference is decided by its meaning and reality, whether we know this or not. Since even this frame depends on our modes of representation, it is less than ‘God’s point of view’.

Characteristic for \((TR^4)\) is the following Thesis:

\[\text{1This idealized notion of frame of reference allows for an idealized notion of verification/assertability: We can postulate that in the best-founded frame of reference each sentence will be decided by making use of Lindenbaum’s Lemma for First Order Theories (that if a theory does not include a proof for a sentence p, there is an extension of that theory including non-p, which is consistent if the original theory was). This guarantees bivalence. So making truth an epistemic concept does not entail intuitionistic logic. What we need is a revisionist notion of negation: That which cannot be deduced or is needed to enhance explanatory power is said to be non-true (Cf. negation in the language PROLOG), so might even be said to be false. More needs to be said here, but the separability of metalogical considerations from epistemology should be clear. This position is revisionistic because some statements – even hopefully so – may correspond to the facts (like “There are exactly the unobservable entities which occur in our theories”), but are said to be not true, because truth requires more than that correspondence!}\]
(Thesis 3) The two aspects of truth in a dual aspect theory cannot be reduced to one another.

Justification:

1. There are statements which might be true in the correspondence sense of truth (like ‘There are exactly the unobservable entities which occur in our theories’), but which by their meaning and the existence of the epistemic gap can never satisfy the justification demand in (TR4).¹
2. On the other hand there might be statements which are part of our best-founded frame of reference but which because of systematic shortcomings of our cognitive faculties do not correspond to reality.²

The internal realists understanding of truth is no idealism. Being a type of realism an even stronger concept of truth can be taken to be meaningful. This is the concept of truth of (R5), leaving aside the justifiability requirement:

(TR5) ‘F (a)’ is true-in-L₁ if and only if the space time area a has the structure F.

Whether we can recognize (absolute) truth does not matter. Truth in this style can be defined for arbitrary languages. The epistemic gap acknowledged by internal realism allows for this ex negativo specified concept of truth: correspondence from an absolute point of view. This external concept can be used by the internal realist to formulate the sceptical possibility.³

Conclusion

The internal realist has, at last, to answer the question how truth can be justified if guaranteed correspondence is impossible because of the epistemic gap. One can appeal here to evolutionary considerations in the context of the internally realistic picture of knowledge. The internalistic truth concept is explanatory: that the orientation on well-founded opinions in our interactions with the world has been successful is explained by their regular dependence on structures of reality. This appeal of a realist to the principle of the best explanation to establish a link between justification and truth has been criticized as being a vicious circle, since it presupposes a link between this very

¹This example statement may well correspond to the facts, We could never – as realists – have it being forced into our best frame of reference, however, since part of that frame is allowing for the epistemic gap. That the statement could be true and we may believe it does not make it true in the sense of (TR4).
²Even if we have no reason to assume that there are such statements, their very possibility undermines the conceptual connection that we would need for a reduction of correspondence to justification.
³As shown by (T) (TR5) does not imply (TR4), and vice versa.
(meta-)justification and truth. Within internal realism, however, this is a
virtuous circle: internal realism starts out with a connection between justifica-
tion and truth and in a further argumentative turn explains why we should
believe in this connection. In the absence of justified doubt all that we take to
be well-founded might be taken to be true and might well be true.

References

The main purpose of this paper is to present a new argument against the Fregean theory of belief reports, proving that if anaphora are devices which keep the same denotation as the expressions they are linked to, then the Fregean theory cannot make sense of a sentence like ‘A believes that she is φ’, where the pronoun ‘she’ works as an anaphor linked to ‘A’.

This paper will be organised into three sections. The first will illustrate the Fregean theory of belief reports and two arguments in favour of this theory. I shall show that neither argument offers a sufficient reason to prefer the Fregean theory to other existing theories of belief reports. The second section will discuss two arguments put forward by Stephen Schiffer (2003) against the Fregean theory. Some successful replies to these will be considered. Finally, the third section will present my new argument against the Fregean theory of belief reports, which takes its cue from Schiffer’s arguments examined in the second section.

I.

In his classic article ‘On Sense and Reference’, Gottlob Frege maintained that every (non-empty) expression has two distinct semantic values: a denotation (or referent) and a sense (or meaning), the latter being a mode of presentation (or concept) of the former. For example, in the sentence (1) the name ‘Fido’ has a denotation (i.e. Fido) and a sense (S_Fido), which is a mode of presentation of Fido. This sense contributes, together with the sense of the predicate ‘is a dog’ (S_is a dog), to the sense of the sentence (1) (i.e. <S_Fido, S_is a dog>), called thought or Fregean proposition.
1. Fido is a dog

This proposition is also what Ralph believes if the sentence (2) is true. According to the Fregean theory of belief reports, in fact, the logical form of (2) is (3), where $B$ is the dyadic relation of belief holding between Ralph and the Fregean proposition $<S_{Fido} \cdot S_{\text{is a dog}}>$.

The construction (3) also reveals that the occurrence of the name ‘Fido’ in (2) does not refer to its customary denotation, Fido, but to its customary sense, $S_{\text{is a dog}}$. The same applies to the occurrence of the predicate ‘is a dog’, which denotes $S_{\text{is a dog}}$ rather than the property of Being a dog.

Ralph believes that Fido is a dog

$B (\text{Ralph}, <S_{Fido} \cdot S_{\text{is a dog}}>)$

The Fregean theory of belief reports is notably supported by two arguments: first, this theory offers a solution to the so-called ‘Cicero’/’Tully’ case; second, it accounts for the intuitive (presumed) invalidity of certain inferences involving belief sentences.¹ I shall show in the present section that neither of these two arguments is compelling.

Let us start by examining the former argument and, to that end, let us consider the following principles/theses:

**Weak Disquotation Principle:** If a subject $S$ at a certain time $t$ in a given possible world $w$ has the disposition to assert/accept sincerely, on reflection and competently the sentence ‘$p$’, which lacks indexical or pronominal devices or ambiguities, then the sentence ‘$S$ believes at $t$ in $w$ that $p$’ is true.

**Thesis on Belief and Referent:** Proper names that co-refer in ordinary linguistic contexts also do so in belief contexts.²

**Principle of Substitutivity salva veritate of co-referential terms:** Given two co-referential terms ‘$a$’ and ‘$b$’, and a sentence ‘$p$’ containing ‘$a$’, the truth-value of ‘$p[a]$’ is the same as the truth-value of ‘$p[b]$’.

**Rationality Principle:** Given a subject $S$, a temporal instant $t$, a possible world $w$, and a sentence ‘$p$’ which lacks indexical or pronominal devices or ambiguities, if the sentences ‘$S$ believes at $t$ in $w$ that $p$’ and ‘$S$ believes at $t$ in $w$ that non-$p$’ are both true, then $S$ is irrational.

Consider also the following sentences:

2. Cicero is bald.
3. Tully is not bald.
4. Tom believes that Cicero is bald.
5. Tom believes that Tully is not bald.
6. Tom believes that Cicero is not bald.

¹By belief sentences I mean sentences of the form ‘$S$ believes that $p$’, where $S$ is a subject and ‘$p$’ is a sentence.
²By belief contexts I mean linguistic contexts of the form ‘believes that …’.
The ‘Cicero’/‘Tully’ case can be presented as a list of four steps preceded by a hypothesis. I call this argument 'C/T'.

Hypothesis: Tom is rational.

Step I: Not realising that the names ‘Cicero’ and ‘Tully’ co-refer, Tom has the disposition to assert sincerely, on reflection and competently (4) and (5).

Step II: The sentences (6) and (7) are true [from Step I using the Weak Disquotation Principle].

Step III: The sentence (8) is true [from Step II – viz. from (7) – assuming the Thesis on Belief and Referent and using the Principle of Substitutivity \textit{salva veritate}].

Step IV: Tom is irrational, against the initial Hypothesis [from Steps II and III – viz. from (6) and (8) – using the Rationality Principle].

This way of presenting the ‘Cicero’/‘Tully’ case enables the determination of its exact import: the set including all principles/theses underlying this case – i.e. the Weak Disquotation Principle, the Thesis on Belief and Referent, the Principle of Substitutivity \textit{salva veritate} and the Rationality Principle – is inconsistent. So, in order to solve the puzzle, at least one of these principles/theses must be rejected.

Proposing a solution, Frege suggested a rejection of the Thesis on Belief and Referent: since – he maintained – the occurrences of the names ‘Cicero’ and ‘Tully’ within belief contexts refer to different senses (instead of referring to the same individual, like in ordinary contexts), the Thesis on Belief and Referent is not met. The Principle of Substitutivity \textit{salva veritate} (which is applicable only to co-referential terms) cannot therefore be applied to replace ‘Tully’ with ‘Cicero’ in (7), i.e. to move from Step II to Step III.

Undeniably, the Fregean solution succeeds in solving the ‘Cicero’/‘Tully’ case. Nevertheless, some alternative solutions to this case are available, consisting in the rejection of the Rationality Principle (as proposed by the Neo-Russellian philosophers Nathan Salmon and Scott Soames) or the Weak Disquotation Principle or the Principle of Substitutivity \textit{salva veritate}. Whereas a rejection of the Principle of Substitutivity \textit{salva veritate} is not particularly attractive because of the great importance of this principle, the strategy of renouncing the Weak Disquotation Principle or the Rationality Principle seems preferable to the one that rejects the Thesis on Belief and Referent: since this thesis follows from the important Thesis of Semantic Innocence, a rejection of the former thesis would entail a rejection of the latter as well.

**Thesis of Semantic Innocence:** The semantic value of an expression does not depend on the linguistic context in which such an expression occurs.

Another critical point against the Fregean theory is that, using the very same principles/theses underlying the ‘Cicero’/‘Tully’ case, it is possible to construct another case, very close to Saul Kripke’s (1979) ‘Londres’/‘London’ case, that the Fregean theory is unable to solve. In order to present such a new case, consider first the following sentences:
7. Londres is pretty.
8. London is not pretty.
9. Pierre believes that Londres is pretty.
10. Pierre believes that London is not pretty.
11. Pierre believes that London is pretty.

Here is the new case in question, presented as a list of four steps preceded by two hypotheses. I call this argument 'P/L'.

**Hypothesis 1:** Pierre is rational.
**Hypothesis 2:** Pierre speaks competently a language that contains all English words plus the name ‘Londres’.

**Step I:** Not realising that the names ‘Londres’ and ‘London’ co-refer, Pierre has the disposition to assert sincerely, on reflection and competently (9) and (10).

**Step II:** The sentences (11) and (12) are true [from Step I using the Weak Disquotation Principle].

**Step III:** The sentence (13) is true [from Step II – viz. from (11) – assuming the Thesis on Belief and Referent and using the Principle of Substitutivity *salva veritate*].

**Step IV:** Pierre is irrational, against the initial Hypothesis [from Steps II and III – viz. from (12) and (13) – using the Rationality Principle].

The Fregean strategy of disallowing the move from Step II to Step III – viz. from (11) to (13) – by renouncing the Thesis on Belief and Referent is inapplicable to the new case. In fact, since the names ‘Londres’ and ‘London’ (as names of different languages) are inter-translatable, then they are synonymous, namely they have the same sense. The occurrences of ‘Londres’ and ‘London’ within belief contexts will therefore be co-referential, coherently with the Thesis on Belief and Referent. As a result, the Principle of Substitutivity *salva veritate* can be successfully applied in (11) to replace ‘Londres’ with ‘London’, i.e. to move from Step II to Step III, leading finally to the absurd conclusion that Pierre is irrational (Step IV). So, because of this new case, the fact that the Fregean theory solves the ‘Cicero’/‘Tully’ case cannot be taken as a compelling argument in its favour.

Let us examine, now, a second (presumed) argument supporting the Fregean theory of belief reports. To that end, consider the following inference:

(Inf1)

(P1a) Tom believes that Cicero is bald.
(P1b) Cicero is Tully.
∴ (C1) Tom believes that Tully is bald.

One problem arises in connection to this inference: on one hand, it seems logically valid; on the other, it seems intuitively invalid in light of the ‘Cicero’/‘Tully’ case.
The thesis that (Inf1) is invalid is endorsed by the Fregean theory. In fact, according to this theory, in the ‘Cicero’/‘Tully’ case the sentences (P1a) and ‘Tom believes that Tully is not bald’ are both true (see C/T Step II). If the latter sentence is true, then (C1) will be false: this easily follows on the basis of the Rationality Principle, according to which the sentences (C1) and ‘Tom believes that Tully is not bald’ cannot be simultaneously true. Once it is stated that (C1) is false, whereas (P1a) and (P1b) are true, the conclusion that (Inf1) is invalid follows.

In this way, the Fregean theory of belief reports solves the aforesaid problem connected to (Inf1) by defending its intuitive (presumed) invalidity. It should be pointed out, on the other hand, that other solutions to such a problem are available. According to Salmon and Soames, for example, the inference (Inf1) is valid. Our misleading intuitions concerning the apparent invalidity of (Inf1) are due simply to confusion between the semantic content of (P1a) and (C1) and their pragmatic content. The semantic content of both these sentences is represented by the linguistic construction (14), where <Cicero, Baldness> is the Russelian proposition composed of the individual Cicero and the property of Baldness. So, the inference (Inf1) will be valid, (P1a) and (C1) having the same semantic content.

12. \[B(\text{Tom}, <\text{Cicero}, \text{Baldness}>)\]

On the contrary, the pragmatic contents of the utterances of (P1a) and (C1) differ. Precisely, the pragmatic content of (P1a) is represented by the linguistic construction (P2a), which can be read as follows: Tom has the disposition to inwardly assent (\(\text{BEL}\)) to the proposition <Cicero, Baldness> when it is presented under the guise (or mode of presentation) corresponding to the sentence ‘Cicero is bald’ (\(\text{MP} <\text{Cicero is bald}>\)). Instead, the pragmatic content of (C1) is represented by (C2), which can be read in this way: Tom has the disposition to inwardly assent to the proposition <Cicero, Baldness> when it is presented under the guise corresponding to the sentence ‘Tully is bald’ (\(\text{MP} <\text{Tully is bald}>\)).

\[\begin{align*}
\text{(Inf2)} \\
\text{(P2a) } \text{BEL} (\text{Tom}, <\text{Cicero}, \text{Baldness}>, \text{MP} <\text{Cicero is bald}>) \\
\text{(P2b) Cicero is Tully} \\
\therefore \text{(C2) } \text{BEL} (\text{Tom}, <\text{Cicero}, \text{Baldness}>, \text{MP} <\text{Tully is bald}>)
\end{align*}\]

In the ‘Cicero’/‘Tully’ case, the linguistic construction (P2a), which represents the pragmatic content of (P1a), is true whereas (C2), representing the pragmatic content of (C1), is false. People who wrongly take the inference (Inf1) as invalid in light of the ‘Cicero’/‘Tully’ case simply mistake (Inf1) with an inference such as (Inf2), the latter being actually invalid. This is the reason of the apparent invalidity of (Inf1), according to Salmon and Soames.

---

\[\text{1Salmon (1986), p. 111.}\]
II.

In the previous section, I showed that neither of the two alleged arguments in favour of the Fregean theory of belief reports – i.e. its solution to the ‘Cicero’/‘Tully’ case and its justification of the intuitive (presumed) invalidity of (Inf1) – is really compelling. Here, I shall discuss some arguments put forward by Stephen Schiffer (2003) against the Fregean theory, particularly against its thesis that the occurrences of the expressions within belief contexts do not preserve their customary denotation.

The first of Schiffer’s arguments I am going to consider (call it Fido argument) is the following:

(i) If the Fregean theory is true, then (a) “Fido” occurs in “Ralph believes that Fido is a dog” as a singular term whose referent is a concept [mode of presentation] of Fido.

(ii) If (a), then the following inference (Inf3) is valid:

(Inf3)

(P3) Ralph believes that Fido is a dog.
∴ (C3) \( \exists x (x \text{is a concept} \& \text{Ralph believes that } x \text{is a dog}) \).

(iii) But the inference isn’t valid; given the truth of the premiss, the conclusion is also true only in the unlikely event that Ralph mistakes a concept for a dog.

(iv) ∴ The Fregean theory is not true.¹

In my opinion, the Fido argument is unsound, because Step iii) is false. In fact, if ab absurdo – as Schiffer claims in Step iii) – the conclusion of (Inf3) was true only if Ralph mistook a concept for a dog, the logical form of the conclusion would be:

13. \( \exists x (x \text{is a concept} \& B(\text{Ralph, } <x, \text{Being a dog}>)) \).

The construction (15) states that Ralph believes a proposition composed of a concept and the property of Being a dog, i.e. he believes that a concept exemplifies such a property. But this is incompatible with the Fregean theory, according to which a proposition cannot contain properties (e.g. Being a dog) but only concepts/senses. In fact, the logical form of the conclusion of (Inf3) is not (15) but:

14. \( \exists x (x \text{is a concept} \& B(\text{Ralph, } <x, S_\text{is a dog}'>)) \).²

¹Schiffer (2003), p. 27. This argument originates with Adam Pautz.
²For reasons different from mine, the unsoundness of Schiffer’s Fido argument is also maintained by Massimiliano Vignolo (2006).
Let us consider, now, another of Schiffer’s arguments against the Fregean theory of belief reports (call it the brother-in-love argument):

There are numerous instances where singular terms occur in that-clauses apparently as referring not to concepts of things but to the things themselves. [...] When your husband’s brother says to you, “I believe that I’m falling in love with you”, isn’t it obvious that both utterances of “I” refer to him [...]?

Although – I think – Schiffer is right in claiming that the two occurrences of ‘I’ refer to only one individual, such a claim is not immediately obvious. In the next section, I shall present a new argument against the Fregean theory of belief reports (call it new brother-in-love argument) which, in a sense, develops Schiffer’s intuition contained in the brother-in-love argument. My new argument will be constructed on the model of Schiffer’s Fido argument but, unlike the latter, it will (hopefully) be sound.

III

The new brother-in-love argument goes as follows:

(i*) If the Fregean theory is true, then (b) the second occurrence of ‘I’ in the sentence ‘I believe that I am falling in love with you’, asserted by your husband’s brother and addressed to you, refers to the concept of your husband’s brother.

(ii*) The following inference (Inf3*) is intuitively valid:

(Inf3*)

(P3*) I believe that I am falling in love with you.
∴ (C3*) There is someone who believes that he (himself) is falling in love with you.

(iii*) The pronoun ‘he’ in the conclusion (C3*) is an anaphor linked to ‘someone’. By definition, anaphora keep the same denotation as the term they are linked to. So, ‘he’ in (C3*) will refer to the same individual who has the belief described in (C3*).

(iv*) Suppose that (P3*) is true. Since (Inf3*) is valid, the conclusion (C3*) will be true as well. So, there will be an individual who is both the denotation of ‘he’ and the subject having the belief described in (C3*). But who could such an individual be? Taking into account that (C3*) follows from (P3*), asserted by your husband’s brother, only two answers are available: either such an individual is your husband’s brother; or he/it is a concept of him. Actually, he cannot be your husband’s brother because, coherently with (b), the pronoun ‘he’ in (C3*) refers to a concept and not to a

1Schiffer (2003), p. 25.
person. On the other hand, the wanted individual is not even a concept of your husband’s brother, since it would be highly implausible (or even senseless) to claim that a concept has a belief – viz. the belief described in (C3*).

\[(v*) \therefore \text{The Fregean theory is not true.}\]

The basic idea of this argument could be stated rather simply. In a sentence like (17), the description ‘your husband’s brother’ and the pronoun ‘he’ cannot have the same denotation, according to the Fregean theory of belief reports: in fact, the description refers to your husband’s brother, whereas ‘he’ refers to a concept of him.

15. Your husband’s brother believes that he (himself) is falling in love with you

On the other hand, ‘he’ in (17) works as an anaphor and, by definition, anaphora keep the same denotation as the term they are linked to. So, ‘your husband’s brother’ and ‘he’ must be co-referential, contrary to what the Fregean theory of belief reports states. This entails that one of the following options must be taken: either rejecting the Fregean theory of belief reports; or rejecting the definition of anaphor. Given the intuitive basis of this definition, the rejecting of the Fregean theory of belief reports seems the best option.

References


n 1965, Noam Chomsky published *Aspects of the Theory of Syntax*, and started a revolution in the field of linguistics, as well as breathing life into philosophical views of the mind and cognition, with roots in the 17th century, long believed dead. The revolution has two parts: first, seek explanation of language in terms of a certain, generally Cartesian, theory of mind; second, treat language as a natural phenomenon, and study it according to the principles of the natural sciences.

Before the publication of *Aspects*, Chomsky had already fired the first shot in this revolution. In his “Review of B. F. Skinner’s *Verbal Behavior*” (Chomsky 1959), he argued that behaviorist studies of human language cannot succeed because they look at the wrong phenomena. In place of behaviorism, Chomsky argued that the proper approach to the study of language must look for a mentalistic explanation of language. Indeed, he convincingly demonstrates that the behaviorist explanations advanced by Skinner all rest upon an implicit and unacknowledged reliance on the very mentalism which behaviorism rejects. The appeal to the mental and the mind as something separate from the body and its behavioral doings foreshadows Chomsky’s later

---

1 I would like to thank Bernard Harrison, Mary Ann Christison, Katherine Matsumoto, John Gordon and Kristin Schaub for their comments on earlier versions of this paper. The paper is the product of research started while on sabbatical from the University of Utah in 2005–06.

2 This has been called many things: “Chomskyan Revolution”, “Cartesian Revolution”, “Cognitive Revolution”. As far as I can tell, Chomsky himself has never described himself as having authored a revolution.

3 At the time, he called this the “generative approach”, and it is often referred to by this name, even today, despite Chomsky’s recent claims to have abandoned it in light of results in linguistic research.
explicit appeals to Descartes.

The mentalism which emerges from the goal of refuting of behaviorism and reviving a generally rationalist picture of mind and language is the focus of this paper. I show that Chomsky’s arguments concerning the nature of knowledge of language entail that he has remained true to his self-avowed Cartesian roots, to the detriment of the other side of the revolution: the naturalization of linguistics, the project of bringing linguistics under the umbrella of the natural sciences. I argue that his conception of the nature of language and our knowledge of it depend upon treating knowledge, and the mental in general, as private in a sense incompatible with the naturalization of linguistics. This has significant implications not only for Chomsky’s understanding of the nature of language, but also for the potential which linguistic research might have for resolving issues in philosophy of language and philosophy of mind. Chomsky believes that linguistics has much to tell philosophy about language and knowledge,¹ thus it is important to uncover any unexpected (and unwanted) consequences of the “revolution”.

It is not my contention that the naturalization project itself creates this dilemma, or that others who study language as a natural object must fall victim to any of these problems. My claim is that in setting up the terms of the “revolution”, Chomsky fails to resolve conflicts between his two aims: mentalism and naturalization. It is this desire to retain Cartesianism, with its mentalism, which commits Chomsky to the view that the mental is logically private.

I do not deny that there are passages in Chomsky’s writing which favor naturalization over Cartesianism. However, there are far more instances in which he makes it clear that Cartesianism is central to his views of language and understanding. For example, consider the following.

...we may say that in appropriate circumstances people think, not their brains, which do not, though their brains provide the mechanisms of thought. I may do long division by a procedure I learned in school, but my brain doesn’t do long division even if it carries out the procedure. Similarly, I myself am not doing long division if I mechanically carry out instructions that are interpreted as the very algorithm I use, responding to inputs in some code in a Searle-style “arithmetic room.” Nothing follows about my brain’s executing an algorithm, in this case of that of translation and understanding. People in certain situations understand a language; my brain no more understands English than my feet take a walk (Chomsky 2000, 113).

One can, of course, insist that all of this is metaphorical, a façon de parler.

¹In The Generative Enterprise Revisited (Chomsky 2004), Chomsky expresses his views on the points of contact between linguistics and philosophy, stating that “the most interesting” point of contact between linguistics and philosophy lies in “… what the study of language has to say about questions of epistemology …”. He goes on to say that “… one can conceive of the study of language being one possible paradigm for the investigation of the nature of knowledge, the nature of human knowledge, and the problems of a priori knowledge” (34).
However, at a certain point, one feels that this looks like an excuse rather than a counter-argument. As Groucho Marx said, “Sometimes a cigar is just a cigar.”

One final note, in his writings it is often difficult to tell when Chomsky is talking about language as the universal phenomenon which explains what it is that all human languages have in common, and language as a set of individual languages: English, Spanish, &c. In what follows, when I use the term “language” with no modifier, it has the former sense. I have tried not to confuse the senses in my appeals to Chomsky’s writings, either in discussing or quoting them.

Privacy

My main thesis is that Chomsky’s view of language commits him to the view that the mental is private in the sense which characterizes Descartes’s views. This is the sense of privacy attacked by Wittgenstein in the Private Language Argument.

One piece of evidence that such a notion of privacy been present in Chomsky from early days is found in The Language of Thought by Jerry Fodor. Here Fodor develops and defends the view that in order to account for language we must posit the existence and availability of a “language of thought”: a language which we never speak, and which we would not recognize if it were spoken to us, but which is necessary if we are to account for language acquisition and use. What is striking about Fodor’s analysis is that he recognizes that a condition for accepting this view is the rejection of Wittgenstein’s arguments against logical privacy.

In a chapter entitled “Why there must be a private language”, Fodor mounts an explicit defense of his views against Wittgensteinian criticisms of “private language”. Whether one finds his reasoning persuasive1 is beside the point. I cite Fodor’s argument to show that one of the most renowned Chomskyans gives an explicit statement of what is entailed by the views of language advanced by Chomsky.

In what follows, it will be helpful to mark a distinction between two senses of the term “private”.

1. **Strong (privateS)**: Logical privacy, the sense of privacy attacked by Wittgenstein and Ryle. In this sense, language requires no public markers or criteria of correctness because any such public markers would, as a matter of logic, be irrelevant to language as a mental phenomenon.

2. **Weak (privateW)**: Contingent privacy. In this sense something is private because it is generally hidden, not readily available for inspection. In connection with language we might say that it is private because it is “in the brain”, and that it doesn’t requiring

---

1In Hanna 1979, I argue against it.
immediate behavioral manifestation, though such manifestation is a good indicator of the presence of language.

Chomsky’s view of knowledge of language commits him to the claim that language is private S, while a treatment of language as a natural phenomenon can only hold that it is private W.\(^1\)

The Revolution

The Chomskyan Revolution has two distinct parts.

1. *Cartesianism linguistics*: This advances the view that the goal of linguist research is explanation, not description, and that this explanation must be given in mental terms. Chomsky appeals to a 17\(^{th}\) century distinction between reason and instinct to support his own views on the difference between human language (the product of reason) and animal communication systems (the product of instinct). According to Chomsky language is

... a natural property of the human mind. But nature does not provide man with an instinctive language, or an instinctive faculty of language ... [man] is not ... compelled to respond in a perfect and specific way. This freedom from instinct and from stimulus control is the basis for what we call “human reason” ... (Chomsky 1966, 58-9).

2. *The naturalization of linguistics*: This holds that language is a natural object, and that linguistics is a natural science; consequently, the study of language should be conducted according to the methods and standards of the natural sciences.

These two principles must be kept firmly in mind in any discussion of Chomsky’s theory of language. This is especially important in connection with his discussion of our knowledge of language. It is here that the tensions inherent in these two sides of the revolution lead to a fatal incoherence in Chomsky’s views on the nature of language and our knowledge of it.

Someone who holds that language is a natural object, to be studied like any natural object, can, of course, maintain a commitment to mentalism by treating the “mental” as an aspect of the physical; such a person might hold that mental states are supervenient on physical states or that they are simply “the physical at a level of abstraction”. This sort of view does not entail privacy in anything other than the contingent sense, privacy W.

There are places where Chomsky talks this way. However, his treatment of the nature of knowledge of language proves that he cannot adopt this strategy without abandoning his Cartesianism, a step which he is unwilling to take.

\(^1\)In fact, this is the only sense of “privacy” in which Fodor might succeed in arguing that there can be a “private language”.

202
Knowledge of Language

Chomsky argues that knowledge of language cannot be understood as a skill, Ryle’s knowledge-how. Instead, for Chomsky, knowledge of language is theoretical knowledge, knowledge-that. As his advocacy of biolinguistics as the proper approach to the study of language becomes stronger, it might seem that he would abandon this view. This has not happened.

Chomsky remains committed to the view that knowledge of language is theoretical knowledge, knowledge-that, as the following passages demonstrate.

Suppose we insist on speaking of knowledge of language as a practical ability to speak and understand. Then normal usage must be revised in numerous cases ... Suppose that Jones takes a public speaking course and improves his ability to speak and understand without any change in his knowledge of English, as we would describe the situation in normal usage. We must now ... say that Jones has improved his ability$_1$ to use his ability$_2$ to speak and understand ... But the two occurrences of “ability” in this description are hardly more than homonyms. Ability$_1$ is ability in the normal sense of the word: it can improve or decline, can be inadequate to determine consequences of knowledge, and so on. Ability$_2$, however, remains stable while our ability to use it changes ... In short, the neologism “ability$_2$” is invested with all the properties of knowledge (Chomsky 1986, 10).

... suppose that Jones, a speaker of some variety of what we call “English”, improves his ability to speak his language by taking a public speaking course, or loses this ability because of an injury or disease (then recovers that ability, say, with a drug). Note that a speaker of “Japanese”, under the same circumstances, would recover Japanese, not English ... In all such cases, something remains constant, some property K, while ability to speak, understand, etc. varies. In ordinary usage, we say that K is knowledge of language; thus, Jones’s knowledge remained constant, while his ability to put his knowledge to use improved, declined, recovered, etc. The account in terms of internal representation of a generative procedure accords with informal usage in this case (Chomsky 2000, 51).1

The structure of these arguments is the same: If someone loses the ability to X, and then recovers it, the renewed capacity to X cannot be grounded in an ability to X because this is the ability which was lost. Therefore, whatever grounds or supports the ability to X must be knowledge. Since, in the case of language, this knowledge is “an internal representation of a generative grammar”, it is “theoretical knowledge”, knowledge-that.

1There is a shift in the sense of the term “language” in this passage. In speaking of “what we call ‘English’”, and later of “Japanese”, Chomsky uses the term in the sense in which it picks out specific human languages; when later he talks about knowledge of language, he is using it in the sense of universal language—what underlies and unites all specific human languages.
This argument is flawed in several respects; however, for my present purposes I focus only on what it reveals about Chomsky’s understanding of the concept of knowledge.¹

The central feature of Chomsky’s argument is that theoretical knowledge is distinguished from skills in terms of behavioral manifestation. The following passage shows that what explains language—knowledge and use of—cannot be behavioral.

Note further that other evidence (say, from autopsy, were enough known about the brain sciences) might lead us to conclude that Smith, who never recovered English ... nevertheless retained his knowledge of English intact after having completely lost his ability to speak and understand (Chomsky 2000, 51).

Skill-based knowledge requires manifestation in the behavior of the “knower”. Since someone in a coma is not required to manifest his knowledge of language, it must follow that knowledge of language cannot be skill-based; therefore, it is theoretical. Following this line of reasoning we must conclude that insofar as one can be said to “know how to dance”, whatever supports this knowledge, must itself be known-that. What my dance course improved is my behavioral manifestation of movements; what Jones recovers or improves is the ability to behave in certain ways. According to Chomsky, what the dance course did not, indeed cannot, improve is any underlying knowledge I have of rhythms, movement, muscles, &c.; likewise, what Jones cannot “recover” or improve is his underlying knowledge of language.

**Cartesianism and Mentalism**

One of Chomsky’s perennial targets is behaviorism. Over time he has broadened his understanding of this to include any and all social, practice-based, or skill-based conceptions of language, along with views of language which make inter-personal communication the focus of concern in the study of language and meaning.

At times, he appears to think that most philosophical conceptions of language fall into this camp, and are therefore mistaken. For example, in Chomsky 2000, a recurrent theme is that philosophers simply do not understand what language really is. In discussing Dummett, Chomsky takes him to task for holding that the fundamental conception of language makes Dutch and German different languages “in which people engage”, consequently language is “learned from others and is constructed by rules which it is part of social custom to follow” (from Dummett, “Comments on Davidson and Hacking”, in Lepore, 473. According to Chomsky, this is the “notion of ‘language’ [which] is essential for philosophical purposes, for the theory of meaning in particular” (Chomsky 2000, 48).

To establish a clear demarcation between his understanding of language as

¹I explore some of the other problems in greater detail in Hanna 2006.
a mental phenomenon and this “philosophical” conception, Chomsky looks to the connection between language and behavior, and concludes that no adequate conception of language can connect it to behavior.

... [t]wo people may share exactly the same knowledge of language but differ markedly in their ability to put this knowledge to use (Chomsky 1986, 9).

Thus, it seems that behavioral manifestation (i.e., speech and understanding) is not an indicator of the underlying knowledge, in short, behavioral manifestation and knowledge are not connected.

This passage is remarkable. It arises from Chomsky’s commitment to the idea that the mental is different in kind from the behavioral, and that the latter cannot explain the former. Explanation must be mental, from which it follows that nothing behavioral can provide the foundation for the mental. Or so goes Chomsky’s thinking. The question is whether this commits Chomsky to holding that language is privateS or merely privateW.

This explains how Chomsky can hold that two people can share “exactly the same knowledge” without any outward indication of this sameness. If knowledge of language is not identifiable by the presence of linguistic behavior, no behavior is necessary for attribution of knowledge. But in order to hold this, knowledge of language must be private. The question is whether privateW is sufficient for this. I argue that it is not, and that Chomsky’s account requires privacyS.

In the cases of Jones and Smith, as well as the two individuals just cited, linguistic behavior cannot function as a criterion of knowledge. This completely severs behavioral manifestation from knowledge. In the case described by Chomsky, Jones knows language, but nothing in his behavior is required to show this, nor could anything in his behavior show it, as the case of the two individuals who possess the same knowledge but have no behavioral indicators of this sameness. With no way to access this knowledge, Chomsky forces us to conclude that it is privateS.

Well, perhaps not. There may be another option. One might identify knowledge with a state of the brain. ¹ This would allow us to say that while language is privateW, and, while it stands in need of no behavioral manifestation, it is not in need of no manifestation whatsoever. In order to be credited with knowledge of language, one must, after all, be in or have the appropriate brain state.

But this is not open to Chomsky. By insisting that we have knowledge of language in an “ordinary” sense (viz., “theoretical knowledge”), and that this of knowledge cannot be identified in any way with skill-based knowledge, he commits himself to bringing knowledge of language into line with the

¹I use the term “brain state” as shorthand for any physiological state, including what Chomsky means when he says that

... [t]he faculty of language can reasonable be regarded as a “language organ” in the sense in which scientists speak of the visual system, or immune system, or circulatory system, as organs of the body (Chomsky 2000, 4).
requirements which hold for any sort of propositional knowledge. To find these while still allowing for the two individuals with identical knowledge, but no behavioral evidence of this sameness, we must push knowledge into the individual and make it private. Anything else threatens his claim that knowledge of language is knowledge in the “ordinary sense”.

There is another reason Chomsky cannot take this path. If one takes seriously the idea that knowledge is a brain-state, any distinction between knowledge-that and knowledge-how becomes one of degree or purpose or classification. By treating knowledge of language a brain state Chomsky would no longer have access to a difference in kind between the two sorts of knowledge. The defender of a skill-based analysis of knowledge of language could argue that, like knowledge-that, knowledge-how is a brain-state. Now the only differences between them would lie in their addresses in the brain, so to speak. On such an account, the challenge to explain how someone could recover his knowledge of language is met saying that recovery of some lost skill at some location, P₁, is explained by the retention of another skill, at P₂, thus avoiding the charge of circularity leveled by Chomsky.

If his argument against the skill-based conception of language is to stand, Chomsky must eschew this move. On his account, as on Descartes’s, there can be no requirement of behavioral manifestation for attribution of theoretical knowledge. Thus, when we look at the two people who “share exactly the same knowledge of language but differ markedly in their ability to put this knowledge to use”, we discover that this is possible only because knowledge is disconnected from both behavior and physical states. One’s knowledge is, and must be, private.

Additional content:

**Mentalism and Naturalization**

The project of naturalization cannot be held responsible for this state of affairs. A reasonable response to these arguments might take the following form. Knowledge-that and knowledge-how are both brain-states, which are distinguished from one another by reference to their respective locations in the brain, or the human system, etc. These two sorts of knowledge play sufficiently different roles in human life that we label the former “mental” and the latter “purely physical” or something else.

The only explanation for Chomsky’s resistance to this approach lies in his commitment to what he called “Cartesian Linguistics”. Here we find the explanation, and with it the root of Chomsky’s present problem. In *Cartesian Linguistics*, he devotes some time to discussions of the difference between instinct and innateness, making clear that human language is not an instinct, which is why he says that it is innate.

Descartes makes a sharp distinction between man and animal, arguing that animal behavior is a matter of instinct and that the perfection and specificity of animal instinct makes it subject to mechanical explanation. A characteristic subsequent view is that there is a gradation of intelligence and that perfection of instinct varies inversely with
intellectual ability ... (CL, 58).

[Referring to Herder] ... language is a natural property of the human mind. But nature does not provide man with an instinctive language, or an instinctive faculty of language, or a faculty of reason of which language is a “reflection.” Man’s fundamental quality is, rather, weakness of instinct, and man is clearly far inferior to animals in strength and certainty of instinct ... “Man does not have an unvaried and narrow sphere of activity, where only one task awaits him” [quoted from Herder]. He is not, in other words, under the control of external stimuli and internal drives and compelled to respond in a perfect and specific way. This freedom from instinct and from stimulus control is the basis for what we call “human reason” ... It is this very weakness of instinct that is man’s natural advantage, that makes him a rational being (Chomsky 1966, 58–59).

In attempting to preserve the difference between human “reason” and animal “instinctive behavior”, Chomsky retains the Cartesian view that human reason is both individual and private. Without the former, behaviorism threatens; without the latter, the distinction between humans and animals becomes one of degree, not the unbridgeable difference in kind envisioned by Cartesians or Chomsky himself. It is difficult to see how this could possibly find a place in any naturalized study of language. The idea that there is a sharp distinction between human language and animal communication systems does not appear to fit with anything the “brain sciences” tell us.

Conclusion

What follow from this? Someone might suggest that while it’s true that Chomsky is confused on these points, it has no impact on his overall programme of reform in the study of language by linguists, and it certainly has no impact on philosophy of language. I believe this is a bit too quick.

First, Chomsky believes that linguistics may give philosophy a “paradigm” for understanding knowledge. If what I have said is correct, it follows that any analysis of knowledge which might emerge from linguistics will be Cartesian, root and branch, because it rests on the assumption that the Cartesian analysis is true.

Second, for the same reason, anything which linguistic research purports to tell us about human understanding or the nature of the human mind will carry with it the assumption that the mental (at least significant parts of it) are private. This view is not only controversial, but there is a significant body of literature showing that it is incoherent. Nothing in what Chomsky writes goes any way towards addressing these arguments, not even the locus classicus, the Private Language Argument.1

1Jerry Fodor addresses this argument, although as indicated in §2, does not succeed in refuting
Third, there are areas of research in linguistics which seem to diverge from the Chomskyan paradigm: for example, sociolinguistics, language acquisition (first and second), and pedagogy. It’s rare to hear a linguist say that they reject the Chomskyan paradigm; in fact, most linguists maintain that Chomskyan theory forms the basis of all research on language. However, if one looks at research techniques and results, along with the apparent conviction that the object of study in these fields is language, not merely some superficial aspect or mere side-effect of language, one might be left with the impression that the theoretical underpinning has no noticeable effect on the research itself.  

In this paper I have shown that Chomsky’s conception of language as a natural object, and his commitment to the treatment of knowledge of language as an instance of knowledge—that are in conflict with one another, and that the tensions in this conflict produce a view of language with is deeply suspect from the standpoint of philosophical (conceptual) analysis. Moreover, many of central tenets of Chomskyan theory do not accord with the actual research practices of linguistics, especially in the more socially and practically oriented sub-fields of linguistics.

References

Hanna, Patricia 1979, “Review of Jerry Fodor’s The Language of Thought”, in Philosophical Investigations, 2, Spring, published under the name “Patricia Johnston.”
Hanna, Patricia 2006, “Language as Community Property: What’s Wrong with Chomsky’s Individualism”, in An Anthology of Philosophical Studies, Athens: ATINER.

Wittgenstein’s arguments against privacy. Chomsky, however, neither addresses the argument nor refers the reader to Fodor.

1 Katherine Matsumoto suggested I distinguish between linguists’s actively rejecting the Chomskyan framework and simply getting on with the business of doing research. The latter does not require much, if any, attention to the implications of the Chomskyan framework. In Hanna 2006, I discuss this phenomenon in connection with research in first language acquisition. I believe that it is not “benign neglect, as the description “getting on with the business of doing their research” might suggest; however, a full discussion of this disconnect would require a separate paper.
Lotteries, Knowledge and Vagueness

Julien Murzi, University of Sheffield, UK

Consider the following two dialogues:

Case 1

John (in his office, early afternoon): I know my car is in the driveway. I parked it there this morning.
Mary: So you know that your car has not been stolen?
John: Well, I don't know that. So I guess I don't know that my car is in the driveway.

Case 2

(Friday afternoon. There’s a long queue in the bank)
John: Let’s deposit our cheques tomorrow. I know the bank will be open then—it was open on Saturday two weeks ago.
Mary: So you know that it hasn’t changed its opening times in between? (Banks often change their opening times. The bank might well be closed tomorrow. And we must deposit our cheques before Monday!)
John: You’re right: I guess I don't know that the bank will be open tomorrow. Let’s follow the line and deposit our cheques now.

Both dialogues start with two intuitively true sentences:

1. I know my car is in the driveway (I parked it there this morning).
2. I know the bank will be open on Saturday (it was open on a Saturday two weeks ago).
In Case 1 it is further assumed that John hasn’t checked whether his car is still in the driveway. Similarly, in Case 2 it is further assumed that he doesn't know whether the bank has changed its opening times in between. On these assumptions, there is a strong temptation to attribute the value ‘true’ to the followings:

3. John doesn’t know that his car hasn't been stolen
4. John doesn’t know that the bank hasn't changed its opening times.

If (3) and (4) are true, however, (1) and (2) must be false. For if John knows that his car is in the driveway, then he must know that it hasn't been stolen. Similarly, if John knows that the bank will be open on Saturday on the basis that it was open on a Saturday two weeks ago, he must know that it hasn't changed its opening times in between. Call the intuitively known proposition (e.g. “My car is in the driveway”) an ordinary proposition (henceforth OP). And call the intuitively unknown proposition (e.g. “It hasn’t been stolen”) a lottery proposition (henceforth LP). Then, the following holds:

5. (5) OP → LP

If knowledge is closed under known entailment, i.e. if closure holds:

(CL)  K\(\phi\), K(\(\phi \rightarrow \psi\)) |- K\(\psi\)

we are confronted with a puzzle: (1) and (3) cannot be both true at the same time; nor can (2) and (4). Something must go. But what?

**Contextualist Solutions of the Lottery Paradoxes**

Broadly contextualist theories of knowledge attributions aim at resolving the tension between our initial assumptions—K(OP) and ~K(LP)—by discharging the tacit assumption that knowledge attributions’ truth-values are dependent upon context-invariant epistemic standards. The idea is that contexts come equipped with epistemic standards that may vary from context to context. In some contexts, epistemic standards may be more demanding; in some other, they can be quite lax.\(^2\) Call this common assumption *Epistemic Standards Sensitivity*:

(ESS) Knowledge attributions are semantically subject to contextually dependent epistemic standards.

ESS can be cashed out in a variety of ways. We consider two.

---

\(^1\)As usual, K\(\phi\) reads ‘someone at some time knows that \(\phi\)’.

\(^2\)There is disagreement among Contextualists themselves about the exact nature of epistemic standards. I shall not take up this issue here.
Indexical Contextualism (henceforth IC) is the view that the word ‘know’ belongs to the same linguistic category as indexical expressions such as ‘I’ or ‘now’. The theory has it that knowledge attributions of the form ‘S knows that p’ really express propositions of the form ‘S knows with respect to s that p’, where s is the epistemic standard of the utterance context. Epistemic standards contribute to the propositions expressed—much in the same way as speaker and utterance time contribute to the content expressed by

6. I am hungry now.

Let ‘x know_{LS}\phi’ and ‘x know_{HS}\phi’ express the relations ‘x knows \phi with respect to a low standard s’ and ‘x knows \phi with respect to a high standard s*’ respectively (where s ≠ s*). Then, the theory has it that John can consistently know_{LS} and fail to know_{HS} the very same proposition, owing to the simple fact that ‘know_{LS}’ and ‘know_{HS}’ express different relations. As a result, neither (1) and (3) nor (2) and (4) result in contradiction—anymore than (6) as uttered by me and its negation as uttered by you do.

Proponents of ESS need not treat ‘know’ as an indexical, however. According to Jason Stanley (2005), it is a mistake to let knowledge attributions depend on the attributor’s epistemic standards. Rather, in his view, epistemic standards are to be determined by the practical interests of the subject to which knowledge is attributed. Let p be our target proposition, and let S be the subject in question. Then, if the truth of p has a practical import for S, the epistemic standards will be rather high. If the truth of the target proposition doesn’t make a practical difference for S, they may become quite low. The resulting theory—Subject-Sensitive Invariantism, henceforth SSI—predicts that John and Mary know or fail to know whether the bank will be open on Saturday depending on how much is at stake for them. If, for some reason, they must deposit their cheques before Monday, then they don’t know that the bank will be open on Saturday. If there is no reason for John and Mary to urgently deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday. If, for some reason, they must deposit their cheques before Monday, then they do know that the bank will be open on Saturday.

Unwelcome Consequences

Broadly contextualist theories of knowledge attributions come with a price.

---

1IC has been proposed and defended by, among others, Stewart Cohen, Keith De Rose and David Lewis. See Cohen (1998), DeRose (1992) and Lewis (1996).
As their own proponents often acknowledge, they all allow for uses of `know' whose correctness appears to be ultimately implausible.

Consider first IC. Such a theory makes the content of knowledge attributions dependent on the epistemic standards of the utterance context. But this doesn’t fit very well with the linguistic data. Consider the following dialogue:

**Case 3**
John (in Low Standards): I know the bank will be open on Saturday.
Mary (in High Standards): What you said is true, but you don’t know that the bank will be open on Saturday. (?!)
John: I guess you’re right. (?!)

Case 3 is indeed quite odd. Mary’s statement and John’s answer are in sharp contrast with our intuition that John and Mary are disagreeing one with each other. On the face of it, IC predicts that subjects should not withdraw knowledge claims made in low standards contexts, when the epistemic standards are raised. For why should they, if what is asserted in low standard contexts is different form what is negated in high standards ones? As Marc Richard points out, the theory fails to account for many of our epistemic disagreements.¹

Secondly, IC has problems with modal embeddings. In fact, it countenances dialogues such as the following:

**Case 4**
John (in Low Standards): I know the bank will be open on Saturday (I was there on a Saturday two weeks ago).
Mary: So you know that the bank did not change its opening hours in between.
John: Well, if you hadn’t asked me this question, I would have known that the bank did not change its opening hours. But now that you’ve asked me, I don’t know it anymore.

Similar dialogues are quite problematic. Knowledge that φ, it seems, cannot depend on whether someone asks us whether we know that, say, ψ.

Unsurprisingly, SSI runs into similar problems. Consider the following situation. John left two soccer balls in his car—b₁ and b₂. The two balls are very similar (neither is brand new and both are quite cheap), except from the fact that, if John doesn’t bring back b₁ to Jones, Jones will kill him. Now suppose John is in his office—he left his car in the driveway. Then, according to SSI, John knows that b₂ is in the car but fails to know where b₁ is.

Semantic Blindness

Both IC and SSI require, it would seem, that subjects are (at least to some degree) systematically blind to the semantics of the language that they speak. This is something that theorists committed to ESS are willing to accept. Thus Keith De Rose:

Contextualism about “knows” doesn't fare […] well, I freely admit. So I suppose there must be something to the charge that if contextualism is true, we suffer from some degree of ‘semantic blindness’: speakers are to some extent blind to the context-sensitivity of ‘knows’ (De Rose 2007).

The appeal to semantic blindness, however, is not a very attractive move: too many theories would become acceptable, if one was freely allowed to ignore their costs.

ESS-theorists typically react by stressing that this is no objection against ESS, since—they claim—every non-skeptical epistemological theory requires a certain amount of semantic blindness. Here’s how De Rose ends the passage cited above:

But this is not a consideration that favours invariantism over contextualism, for speakers seem about equally afflicted by semantic blindness whether contextualism or invariantism is correct (Ibid).

In a similar spirit, Stanley writes:

It is not clear that the unintuitive consequences of SSI are any worse than the unintuitive consequences of other standard epistemological theories (Stanley, 2005, p. 131).

But that a viable answer to the problem we started with? We suggest that it isn’t. On this assumption (and on the assumption that CL is not up for grabs), we are left with no choice but to discharge ESS.

Low Standards Invariantism

So far, only two non-contextualist accounts of knowledge attributions have been seriously considered in the literature: Low Standards Invariantism (henceforth LSI) and Skepticism. Skepticism is the claim that all, or nearly all, our knowledge attributions are just plainly false. Here we shall dismiss it without argument. LSI has it that the epistemic standards to which knowledge attributions respond are invariant across contexts and typically low.1 It is to this latter theory that we now turn.

According to LSI, John knows both the ordinary and the lottery propositions in the examples from which we started. Since, however, we intuitively tend to deny such a knowledge to John—at least as soon as we realize that he doesn’t know the relevant lottery propositions—Low Standards Invariantists have to explain why John has knowledge, in spite of our intuition that he doesn’t. Low Standards Invariantists have provided at least three different answers to the problem.

Proponents of a warranted assertibility manoeuvre, such as Jessica Brown and Tim Black, suggest that in so-called High Standards scenarios the truth-conditions of knowledge attributions are typically satisfied, although their assertibility conditions are not. The idea seems to be that knowledge does not require that one’s epistemic position is strong enough to rule out all the counterpossibilities that are considered in a given context, although asserting that one knows pragmatically conveys the information that all such counterpossibilities have been eliminated. Hence—the idea is—knowledge claims may be infelicitous, or unwarranted, but nevertheless true.

Low Standards Invariantists such as Williamson (2005) and Adler (2006) have suggested a psychological solution to the problem. As a psychological matter of fact, they argue, when epistemic counterpossibilities are waved, we start to lose confidence in our own beliefs. When John considers the possibility that his car has been stolen, for instance, his confidence in his belief that his car is in the driveway is shaken. Eventually, he may become persuaded that he doesn’t know that his car has not been stolen. Yet, Low Standard Invariantists insist, John still knows where his car is. He knows, but he doesn’t know the he knows—and neither do we. His—and indeed our—intuition to the contrary, it is claimed, is owed to mere psychological (as opposed to genuinely epistemic) reasons.

Finally, in a recent paper Adler (2006) has recently defended a more epistemic line of thought. Counterpossibilities, Adler suggests, can undermine knowledge, but only if they enjoin a withdrawal of the corresponding belief. For instance, the possibility that John's car has been stolen can undermine his knowledge that his car is in the driveway only if he ceases to believe this latter proposition. Typical cases advocated by contextualist theorists, however, all have it that the subject in question still holds the relevant belief. Consider again Case 1. John still believes that his car is in the driveway, irrespective of Mary’s contention that he doesn’t know that it hasn’t been stolen. Hence, Adler concludes, what Mary says does not provide epistemic reasons, for John, to withdraw (1).

In whatever version, LSI faces the following problem. The theory first assumes that we typically have knowledge in high standards contexts. Then, it attempts to explain away—in various ways—our intuition that, in such contexts, we don’t have knowledge. This assumption commits defenders of LSI to dogmatism. From the mere fact that we (may) tend to attribute knowledge of the target proposition to John in contexts in which some counterpossibilities are not being considered, it does not follow that John actually has knowledge. Consider the following argument. In order to know that his car is in the

---

1 See Brown (2005) and Black (2005).
driveway at some time t, John should have evidence that his car is in the driveway at t. This requirement rests on the following (plausible) principle:

\[(KE) \text{ One knows that } p \text{ only if one has evidence for } p.\]

Now let \(t^*\) be the time of utterance in Case 1. Then, ex hypothesis, John doesn’t have any evidence that his car has not been stolen at \(t^*\). Therefore, if KE holds, John doesn’t know that his car is in the driveway at \(t^*\) (of course, John is justified in believing that is car his in the driveway at t. But it is no mystery that one can justifiably believe a proposition without knowing it). In order to insist that John nevertheless knows the whereabouts of his car, Low Standard Invariantists would have to deny KE, which would be tantamount to embracing some form of dogmatism.

Of course, this is not to say that it is impossible for us to think that we don’t know things that we actually know. Suppose that, at a given time t, I do remember my credit card number. Suppose also that, at some later time \(t^*\), someone keeps saying that I might have actually forgotten it. I then start having doubts about the correctness of the number I have in mind. Suppose I eventually cease to believe that the number I remember is the right one. If belief is a necessary requirement for knowledge, my knowledge would then be destroyed for merely psychological reasons. However, this case seems quite different from the driveway one. In the credit card case, I still know my credit card number, even though I don't know that I know it. In the driveway case, John never has evidence that his car is in the driveway—on reflection, he is even willing to admit that he doesn't. We thus have a \(\neg KK\phi\) combination in the first case and a \(K\neg K\phi\) combination in the second one. Arguably, the latter cannot be reduced to the former.

**High Standards Invariantism**

I shall argue that both contextualist theories and LSI rest on a common mistake. I will thereby defend a novel non-contextualist solution to the Lottery Paradoxes—one that can be seen as an intermediate position between LSI and full-blooded Skepticism.

Contextualist theorists interpret cases 1 and 2 as providing evidence for their central tenet—ESS. It seems, however, that such cases can be given a more straightforward reading. All that linguistic data show is that speakers tend to systematically withdraw some knowledge claims if properly challenged. Speakers tend to say, for instance, that they know that their car is parked in the driveway, but they will almost invariantly withdraw such a claim if asked whether they know that it hasn’t been stolen. What they really know, they will say, is that their car was in the driveway when they parked it. A natural explanation for a similar phenomenon is that speakers might tend to credit themselves more knowledge than they actually have. Let us call a similar hypothesis High Standards Invariantism (henceforth, HSI). In a nutshell, HSI can be thus stated:
(HSI) \( \forall p \forall S, 'S knows that p' takes the value 'true only if \( \forall q, if p \implies q \) and \( if q \) is not a skeptical claim, then \( S \) is in a position to know that \( q \) (given the evidence she already has).

According to HSI, intuitively unknown skeptical consequences of our target propositions do not undermine knowledge—HSI takes them to be known; intuitively unknown non-skeptical consequences do.

Here is some prima facie evidence for HSI. When we do know, we don’t withdraw our knowledge claims. Consider the following case:

**Case 5**

John (in the driveway, early afternoon): I know my car is in the driveway—it is here in front of me.

Mary: So you know that your car has not been stolen!

John: Of course I know that.

In more controversial cases, however, things change. Consider again our initial examples. John initially asserts (1) and (3), but he is happy to withdraw such claims once that Mary has pointed out to him that (2) and (4) hold. HSI precisely predicts this. Sometimes we know, the theory says, sometimes we don’t. In order to know \( \phi \), one must be in an epistemic position that is strong enough to rule out those non-skeptic epistemic possibilities whose obtaining would undermine \( \phi \). But this just seems to be a fair requirement. Two theses suggests themselves. First, there is only one kind of epistemic standards—the high ones. Second, HSI does not really require semantic blindness. For John is always disposed to acknowledge that he didn’t know in the first place what he claimed to know. Hence, he cannot be properly said to be blind to the semantics of his own language—even though he sometimes credits himself knowledge that he doesn’t have.

If all of this is correct, then we are in a position to say what's wrong with both contextualist theories and LSI. Both kind of theories credit knowledge to subjects whose epistemic position is not strong enough to rule out all the potential counterpossibilities whose actual obtaining would undermine the truth of the target proposition. Since this is in contrast with our intuitions, it is no mystery that both contextualist theories and LSI have unintuitive consequences.

**Skepticism and Vagueness**

At least two different kinds of considerations tell against HSI. Firstly, the theory seems to be at odds with two widely shared normative principles: that knowledge is the norm of assertion and that knowledge is a norm of practical reasoning.\(^1\) Secondly, it may be argued that HSI is likely to undermine most of our knowledge. Here I shall exclusively focus on the second worry.\(^2\)

---


\(^2\) For a defence of LSI from the objections from the knowledge accounts of assertion and of
HSI requires that there be contexts in which skeptical possibilities are properly ignored, though real counterpossibilities are not. Two potential concerns with this proposal are worth considering. First of all, one might question the legitimacy of leaving skeptical counterpossibilities aside when it comes to spell out the truth-conditions of knowledge attributions. Second, and more importantly, it might be reasonably argued that—unless a principled distinction between skeptical and non-skeptical claims is at hand—HSI inevitably collapses on Skepticism.

A moment’s reflection shows that the first worry may be easily addressed. Ordinary speakers typically ignore skeptical possibilities. They are just like Hume at the pub: they are not worried by skeptical fantasies. Therefore, as far as we aim at giving a semantics for ordinary uses of ‘know’, it seems perfectly correct to (i) let knowledge attributions be independent from skeptical counterpossibilities and (ii) assume that skeptical counter possibilities are always known. The second worry is indeed more challenging. How could the proponent of HSI deal with it? Here is one possible way to go.

Consider the question whether I know that I am a student at the University of Sheffield. I am traveling, and it has been a couple of days I haven't checked the University's website. For all I know, the University might have been destroyed, or it might even have ceased to exist as a legal institution. Since I don't know that a similar scenario does not obtain, if CL holds, I don’t know that I am a student at the University of Sheffield either. Now here is the interesting question. Is that a skeptical scenario? This question does not seem to admit a definite answer. The sentence “The University of Sheffield might have ceased to exist in the last few hours”, however implausible it might sound, is surely both metaphysically and practically possible. Yet, at the same time, we feel strongly inclined to consider it as a skeptical claim. The natural suggestion, here, would then be to treat similar cases as borderline cases of the predicate ‘ϕ is a skeptical claim’. If this is correct, the proponent of HSI might then suggest the following package. The very notion of a skeptical claim is vague. Indeed, it is not difficult to build sorites series starting with truly skeptical scenarios and ending with a perfectly mundane possibility. As a result, given HSI, ‘know’ is itself vague, and there will be cases in which it not clear whether we have knowledge or not. Specifically, there will be cases in which our target proposition may entail some proposition that is not definitely skeptical and not definitely non-skeptical. Let’s consider two variations of Case 1:

**Case 1a**
John knows that his car is in the driveway.
(He parked it there one minute ago. Surely no one could have stolen it in such a short time.)

**Case 2a**
John doesn’t know that his car is in the driveway.
An Anthology of Philosophical Studies Volume 2

(He parked it there one week ago. It should be there, but, these days, you never know.)

Neither case constitutes a problem. But now consider:

**Case 3a**
John knows that his car is in the driveway.
(He parked it there this morning, ten minutes ago.)

Does now John know that his car is in the driveway? Well, it depends. On the one hand, it looks as though it is very unlikely that the car had been stolen in such a short time. On the other hand, cars are sometimes quickly stolen in John’s neighborhood. We thus hesitate between attributing to John knowledge of the ordinary proposition (the car is in the driveway) and the lottery one (it is not the case that it hasn’t been stolen). Proponents of ESS hold that conflicting verdicts, here, are due to shifts of epistemic standards. But this doesn't seem correct. It rather looks like, when it comes to similar cases, we hesitate in assigning a truth-value to the claim that the ordinary proposition is known and to the claim that the lottery one is unknown in very the same context. A much more plausible hypothesis, therefore, would be that of including such intermediate cases within the borderline area of ‘know’. In such cases, the relevant instance of HSI will either have an indeterminate truth-value, or a truth-value of a certain degree, or an unknowable truth-value, depending on the theory of vagueness one is adopting. But in neither case do we have a Paradox: the borderliness of the claim that the ordinary proposition is known will always go hand in hand with the borderliness of the claim that the lottery one is unknown.

**Conclusion**

Does HSI fare any better than its contextualist rivals? I suggest that there might be reasons to answer in the affirmative. First of all, it respects our invariantist intuitions. Second, unwelcome consequences are avoided. Third, HSI allows Lottery Paradoxes to be accounted for in one of the three following ways:

The relevant knowledge claim is withdrawn because of psychological phenomena (both the ordinary proposition and the lottery one are known;
We have credited ourselves knowledge that we actually don't possess (both the ordinary proposition and the lottery one are unknown).
We are facing a borderline case of ‘know’ (whichever account of borderliness is adopted, the truth-value of knowledge claims concerning, respectively, the ordinary proposition and the lottery one are the same).

Possibilities (i) through (iii) are jointly exhaustive. Since, in all cases,
knowledge claims concerning, respectively, the ordinary proposition and the lottery one always have the same truth-value, no puzzle arises.

As a result, we are left with a quite austere notion of knowledge. Whenever rational, non-skeptical doubts can be raised, knowledge cannot obtain. But this isn’t obviously an unwelcome result. Knowledge is peculiarly valuable because it is more than mere justified, or probably true, belief. The paradoxes we started with typically arise whenever we credit ourselves knowledge that so and so at a given time t, where our evidence in fact only supports knowledge that so and so at some t’ before t. In this respect, HSI simply reflects the fact that the future is mostly unknown to us—no matters whether it is metaphysically open or not. The supposed evidence for ESS ultimately rests on psychological phenomena that do not have semantic import. However, psychological phenomena should not obscure the semantic truth.

References

De Rose, K.: forthcoming, Bamboozled by our own words: Semantic blindness and some arguments against contextualism.
Abstraction and the Interpretation of Quality

Donald V. Poochigian, University of North Dakota, USA

Experience composes quality and abstraction. Quality exhibits the character of an irrational number, which is resolved nominally by an abstract identity. Quality is simply defined or assumed or understood some way, when definition, assumption, and understanding are wholly unlike what is defined or assumed or understood. These things and others like them are abstractions, evidenced by their being unobservable alone and unobservably affecting what is defined, assumed, or understood. Unobservable alone, however, what is an abstraction? And unobservably affecting quality, how does abstraction interpret quality?

Quality

Quality is content of sensation, and can be content of imagination. Abstraction is content of imagination. Identified as different in different time and space, quality is intrinsically temporal and spatial. Identified as same in different time and space, abstraction is accidentally temporal and spatial. Complication occurs because limitlessly divisible in time and space, whether quality is one or many is ambiguous. 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.

---


Because quality is extended in time and/or space and abstraction is extended in neither, it would appear no extent of abstraction can constitute quality, and no extent of quality can constitute abstraction. Neither constitutive of the other, they are distinct primitive elements. Composing two mutually irreducible constituents, a problem of relationship is presented. Like the traditional mind/body problem, presented are apparent incommensurables.

Being an endlessly reductive nonrepeating sequence like an irrational number, quality is emergent. Being an indubitable indivisible nonreductive sequence, abstraction is emergent. Both quality and abstraction being emergent, they are inexplicable. As such, they are irrational. Indubitable while so, they are primitive elements. Quality being extended in time and/or space and abstraction extended in neither, no quantity of abstraction can constitute quality, and no quantity of quality can constitute abstraction. Neither reducible to the other, they are distinct primitive elements.

Abstraction

Grouped and ungrouped qualitative elements being sensately indistinguishable, distinguishing them is abstraction. 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. 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 supravenience 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.

Abstraction might be thought to suppose mind, but does not. Same experience can be sensation or imagination. Which is a function of attribution of a hypothetical source, whether mind or body. Unobservable such a source is an abstraction. Determining mind as well as body, abstraction presupposes neither. Thus, Hilary Putnam and Richard Rorty ‘have argued, in effect, that a metaphysical view based on a distinction between appearance and an underlying reality is seriously defective and that a purely ‘scientific’ metaphysics is really no better than the wilder views of a hundred years ago.’\(^2\)

---


Abstraction and the Interpretation of Quality

Identified as sourced by body, experience is ‘real.’ Identified as sourced by mind, experience is ‘unreal.’ Both these when experience sourced by either is an abstract identity, exhibiting Edmund Husserl’s concept of ‘bracketing’ whereby, the attempt to doubt any object of awareness in respect of it being actually there necessarily conditions a certain suspension (Aufhebung) of the thesis; and it is precisely this that interests us. . . . the thesis undergoes a modification—whilst remaining in itself what it is, we set it as it were “out of action,” we “disconnect it,” “bracket it.” It still remains there like the bracketed in the bracket, like the disconnected outside the connexional system.1

Understanding the phenomenal as mind and the material as body, both are constituent of the same experience, mutually accompanying a common sensation of quality. They are separated only by ‘bracketing,’ while still remaining ‘like the bracketed in the bracket.’ This is explained by their being hypothetical abstractions functioning like propositional attitudes—‘believing that.’ A propositional attitude is a rule postulating existence of something from which experience and behavior occurring deductively follow. Mind and body are irreducible not because of their properties (effects), which are the same, but because they are different indivisible entities following from such rules. As simple primitives, neither can compose the other.

Ontology

Unobservable like theoretical entities, unlike theoretical entities which are substantial in some sense, like physical forces abstraction is insubstantial.2 Also like physical forces abstraction infuses quality without measurably affecting it. 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. Both physical forces and abstractions being ‘‘something’’, they constitute impalpable entities ontologically distinct from quality.3 That physical forces can vary in time and space renders them physical. That abstraction does not vary in time and space renders it metaphysical.

---

2The modern quantum mechanical view of the three fundamental forces (all except gravity) is that particles of matter (fermions) do not directly interact with each other, but rather carry a charge, and exchange virtual particles (gauge bosons), which are the interaction carriers or force mediators. ['Fundamental interaction.' Wikipedia, the free encyclopedia. Available at http://en.wikipedia.org/wiki/Fundamental_interactions.]
Virtual particles’ being ‘force mediators’ are distinct from the forces mediated, being only ‘interaction carriers.’ Thus, ‘Virtual particles are viewed as the quanta which describe fields of the basic force interactions, which cannot be described in terms of real particles.’ ['Virtual particle.' Wikipedia, the free encyclopedia. Available at http://en.wikipedia.org/wiki/Virtual_particle. Italic added.]

As such, abstraction is the fugitive ‘meaning’ identifying a symbol. John Searle hints as much in his ‘Chinese room’ example where behavior of the knowledgeable and ignorant are indistinguishable. Syntactical meaning is incomplete because syntax is a sequence when sequence is ambiguous, identifiable cardinally or ordinally. Sequence is a matter of interpretation, and interpretation is abstract identity.

Not substituting for quality, abstraction does not transmute from quality as Colin McGinn presumes. Such transmutation requires a ‘hidden structure’ constituting an alchemic transformation of substances bordering on the ‘miraculous,’ akin to the Roman Catholic and Eastern Orthodox communion’s transubstantiation of bread and wine into flesh and blood.1 Thus, McGinn’s pronouncement in this respect does not differ fundamentally from the pronouncement of ‘the great mystery which [Jesus Christ] left us as an everlasting covenant,’ or call to ‘Let us proclaim the mystery of faith,’ declared in the Eucharistic Prayer IV of the Roman Catholic Church.2

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 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.

Epistemology

Insensate unlike quality, how is abstraction known? Qualitative experience is sensation, and abstract experience is sense. There is a sensation of quality or sensation like experience of quality concerning dreams, hallucinations, and imagination. There is not a sense of quality or anything like quality. There is a sense of abstraction, but not a sensation of abstraction.

Thinking distinction between sensation and sense forced, consider Wittgenstein himself concerning meaning makes reference to ‘the sense of a sentence,’3 and again, ‘that sentences have the same sense.’4 Speculatively sense is the manner in which G. E. Moore understood good an object of experience. By contrast, concerning “‘cognitively significant sentences,’” logical positivism does not consider sense.5 Neither can it differentiate whole and parts when cognitively indistinguishable. One can have only a sense of the same cognitive experience as whole or parts.

---

2Italics added for emphasis.
3Wittgenstein. Part II., vi, 181e.
4Wittgenstein. nt. 20, 10e.
Abstraction and the Interpretation of Quality

Sensation is a ‘1sense 3: b: a definite but often vague awareness or impression,’\(^1\) which is to ‘2sense GRASP, COMPREHEND.’\(^2\) Manifest is ‘intuition c: the power or faculty of attaining to direct knowledge or cognition without evident rational thought and inference.’\(^3\) It is a primitive act of, ‘apprehension 1 a: the act or power of perceiving or comprehending.’\(^4\) 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 amount.’\(^5\) As so, it constitutes ‘comprehension 1 c: the capacity for understanding fully.’\(^6\)

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.

Resolved are the ‘concealed contradictions’ of which mathematical intuitionists caution.\(^7\) 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,’\(^8\) any quality can uninterruptedly transmute into any other quality without exhibiting a concealed contradiction (or better ‘inconsistency’).\(^9\) Acknowledging this, all constituents of an unbroken continuum between any two recursive constituents are incorporated in endless reduction.

Whether an antinomy or contradiction, then, a concealed inconsistency cannot occur within an infinite series. An antinomy is an impossibility, an element concurrently constituent of incompatible sets. A contradiction is a limit, an element concurrently constituent of compatible sets, ambiguously constituent of one or other or both. An antinomy cannot occur a concealed inconsistency because it cannot occur. A contradiction cannot occur a concealed inconsistency because it is a limit, and a limit cannot occur until identified as such.

Sense of identity resolves how elements are known to be fused or diffused when observationally indistinguishable, identity being self-evident. E. J. Lowe introduces this difficulty whereby, ‘resemblance between particulars is always

---

2v’2sense.’ In: Mish & et. al. 1066.
3‘intuition c.’ In: Mish & et. al. 614.
4‘apprehension 1 a.’ In: Mish & et. al. 57.
5‘comprehend 2.’ In: Mish & et. al. 237.
6v ‘comprehension 1 c.’ In: Mish & et. al. 237.
9 Nagel & Newman. 23.
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."'1 Appropriate specification constituting a rule, 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 by common meaning, which is constancy of abstract identity throughout experiential qualitative transmutation.

**Continuum**

Abstraction identifiable 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.2 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*.

Being self-identifying, abstraction is self-contained. Being self-contained, abstraction can be related to other abstraction only by identity as sequenced, this constituting an encompassing abstraction. 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.

Unlimited, continua are irrational; limited, they are rational. Continua are rational when abstractly limited, which is nominal. Continua are rational, then, only nominally.

---

Limit

Christian thought incorporates the nominal character of continua by introducing the Will of God. Enlightenment thought excludes the nominal character of continua by introducing the ‘consecutive digits repeating themselves indefinitely’\(^1\) of Georg Cantor’s set theory in mathematics, and ‘The energy of the universe is constant while entropy strives toward maximization’\(^2\) of Rudolf Clausius’ entropy theory in physics. Integrating them into the foundation of modern science is Charles Saunders Peirce.

Peirce presents experience as a commutative ring with unity.\(^3\) As Ian Stewart notes, though, “commutative ring with unity. . . . is a trifle cumbersome,” and so by convention is simplified to “A ring.”\(^4\) Exhibited is a function whereby constituents reciprocally fuse into the same indistinguishable whole by repetitive iterative application of the ‘+’ conjunctive function in any sequential order, and diffuse into the same distinguishable parts by repetitive iterative application of the ‘.’ disjunctive function in any sequential order.\(^5\) Constituents as fused whole are an object, and as diffused parts are not an object, constituting a self-generating system.

Presuming infinite not endless divisibility of quality Peirce concludes, All phenomena therefore are inferences or interpretations, and so are to some degree conceptual. Combining this fact with the apparent regress involved in predication, Peirce drew the conclusion that all cognitions are the results of inference, and so are logically determined by prior cognitions of the same object. The fact that this involves an infinite regress Peirce accepted once he was convinced that the regress was not vicious.\(^6\)

Drawing from Cantor’s set theory,\(^7\) ‘the regress’ is judged ‘not vicious’ because assuming,

\(^1\)‘irrational number.’ In: Mish & et. al. 619.
\(^3\)‘In a commutative ring with unity, every maximal ideal is a prime ideal.’ [‘Maximal ideal, Properties.’ Wikipedia, the free encyclopedia. Wikimedia Foundation, Inc. Available at http://en.wikipedia.org/wiki/Maximal_ideal. [Last modified 04:57, 26 April 2007]] By clarification, If R is a commutative ring, then an ideal P of R is prime if it has the following two properties:
- whenever \(a, b\) are two elements of \(R\) such that their product \(ab\) lies in \(P\), then \(a\) is in \(P\) or \(b\) is in \(P\).
- \(P\) is not equal to the whole ring \(R\). [‘prime ideal.’ Answers.com. Wikipedia. Available at http://www.answers.com/topic/prime-ideal.]
\(^5\)Stewart. 78.
\(^7\)‘The Cantor set, introduced by German mathematician Georg Cantor, is a construction of a set of points lying on a single line segment, and involving only the real numbers between zero and one.’ [Farlex. ‘Cantor set.’ The Free Dictionary. Available at http://encyclopedia.thefreedictionary.com/Cantor+set.] Relevantly the Cantor set involves ‘only the real numbers between
The dichotomy paradox

“You cannot even start.”

“That which is in locomotion must arrive at the half-way stage before it arrives at the goal.” (Aristotle Physics VI:9, 239b10),

whereby a ‘goal’ or limit is incorporated. Although quality may endlessly divide at any level of scale, as levels of scale descend, all divisions converge ‘in locomotion’ on the limit.

Quality exhibits no ‘goal’ or limit, though, not only endlessly dividing within any level of scale, but in levels of scale. Now all divisions tend to disperse, converging only accidentally, not necessarily, rendering ‘the regress . . . vicious.’ A rational expansion either terminating or becoming periodic—a form of termination returning to an origin—quality is irrational because neither terminating nor becoming periodic except accidently. Since, ‘there are . . . many more possible disordered states than ordered ones . . . a dynamically ordered state . . . is thus “the most improbable case conceivable . . . an infinitely improbable configuration of energy,”’ there is no entropic state of maximum disorder.

Illustratively, considering quality, Jonathan Edwards observes if a body placed between two other bodies is divisible, and ‘if it be so that the parts can be broken still finer and finer, they can be broken so fast as not to retard the motion of the [two other bodies] at all.’ Endlessly being ‘broken still finer and finer,’ quality presents a palpable something made of nothing, rendering it irrational. Represented by a reductive sequence to the right of a decimal composing ‘a number that can be expressed as an infinite decimal with no set of consecutive digits repeating itself indefinitely and that cannot be expressed as a quotient of two integers,’ mathematically is constituted an irrational number.

Incorporated into physics, a limit constitutes entropy whereby generalized requires ‘the heat death of the universe.’ Disputing necessity of stasis, employing the ‘substitution effect,’ Paul Samuelson argues economically, ‘The tendency toward diminishing returns has just been offset by the technical shift.’ Accepting the law of conservation of energy, this occurs by serial zero and one,’ whereby the unit interval [0, 1] introduces set limits. [‘Smith-Volterra-Cantor set.’ The Free Dictionary. Available at http://encyclopedia.thefreedictionary.com/Smith-Volterr- Cantor+set].


4 Flower & Murphey. 143.

5 ‘irrational number.’ Mish, et. al. 619.


8 Samuelson. 747.
substitution of self-identifying entropic technology with self-identifying entropic technology in endless transitive unentropic sequence. Linear rather than circular, this analysis is otherwise similar to that of the Buddhist Wheel of Life.

Neither is it inconsistent with Jonathan Edwards’ analysis of quality as unlimited. Utilizing Edwards’ ‘parts . . . broken still finer and finer,’ resources are understandable as increasing as much as decreasing. Unlimited, however, analysis is impossible, calculation having no product. Analysis is possible only by introducing a nominal limit. Nominal, such a limit is an abstraction. Analysis requiring a nominal limit renders an objective universal account impossible.

Nominal

This nominal character is exhibited in three elemental combinations of continuum and limit. A continuum can be limited without and within, within, or unlimited. Constituted respectively is a line, a ring, and a field.

All continua are generated by analogical likeness to an archetype. A quality being endlessly divisible, it can be identified as constituent of a class at a level of scale and not another. Practically a rule cannot be complete at all levels of scale.1 Practically an archetype is complete at all levels of scale. Therefore, a qualitative continuum can be generated by an archetype, not a rule. Archetypes need not be real. Therefore, a qualitative continuum is generated by an archetype, real or imagined.

Identified by a line is a particular constituent indistinguishable from a continuum constitutive of implication. Identified by a ring is a particular constituent of a continuum constitutive of conjunction. Identified by a field is a particular distinguished from a continuum constitutive of tautology.

Line

Lines and rings are iteratively generated by likeness to an inconstant analogical archetype, constituents generated by the same archetype constituting a subset of the linear set. Each such archetype constitutes an internal limit of the linear set, concurrently constituent of the set and subset. Distinguishing a line is irreversibility of archetypal identifiers, a sequence analogically generated in one order not generated in reverse order. What seems a terminus of preceding elements does not seem a terminus of subsequent elements.

Determining a set limit is constituency in more than one set. Exhibited is Ernst Zermelo’s “axiom of choice” stipulating, ‘for any set whose members are sets that are non-empty and mutually exclusive, there exists at least one set having exactly one element in common with each of the sets belonging to the

---

An Anthology of Philosophical Studies Volume 2

original set.'¹ Such an element constitutes ambiguity, concurrently constituent of two or more ‘non-empty and mutually exclusive’ sets.

Distinguishing rings are algebraic unknown numbers, constituting resolution of their disjunctive limits. Being arbitrary employment of exclusive disjunction endlessly, an unknown number need not be generated to know it is complete. Unnecessary, identity of lines is nominal. Any apparent contradiction is resolved casuistically by distinguishing some uncommon characteristic.

An unknown number is formally indistinguishable from an irrational number, except progressing to the left rather than regressing to the right of a decimal. Although substantively rational, an unknown number is indeterminate like an irrational number. Identified are all resolutions of Bertrand Russell’s ‘Between two perceived perspectives which are similar, we can imagine a whole series of other perspectives, some at least unperceived.’² Endless in extent, generated is an unascertainable numeral.

Like an irrational number, an unknown number is indeterminate, but unlike an irrational number, an unknown number is additive from identifiable constituents.³ Endlessly subtractive, an irrational number has no identifiable constituents from which it is composed, rendering it irrationally incomprehensible. Endlessly additive, an unknown number has identifiable constituents from which it is composed, rendering it rationally comprehensible.

Ring

Considering a ring, exhibited is a function whereby constituents reciprocally fuse into the same indistinguishable whole by repetitive iterative application of the “+” conjunctive function in any sequential order, and diffuse into the same distinguishable parts by repetitive iterative application of the “.” disjunctive function in any sequential order.⁴ Constituents as fused whole are an object, and as diffused parts are not an object.

A ring can be a commutative unity constituted by recursive identity of the sequence with a correspondent sequence, or a corecursive hyperset constituted by corecursive identity of each element with its opposite in marginal ordinal order counting from the opposite limit. As corecursive, a commutative ring is self-identifying; as not corecursive, a commutative ring with unity is not self-identifying. Being not self-identifying, a commutative ring with unity is indeterminate. Any rule identifying it, if not corecursive, is identifiable by

³‘In algebra, a letter that represents any number is call a literal number.’ [Hart, W. W., Schult, V. & Swain, H. (1957). First Year Algebra. Boston: D. C. Heath and Company. 9] In resolving the translation of a ‘literal number’ into a number, ‘The problem is to find what value this literal number must have so that the two sides of the equation will have the same numerical value. For this reason, the literal number in the equation is called the unknown number.’ [Hart & et. al. 42] See Derbyshire, J. (2006). Unknown Quantity: A Real and Imaginary History of Algebra. Washington, D. C.: Joseph Henry Press.
⁴Stewart, I. 76-78.
induction from elements or deduction from a rule. These or this not being corecursive, is identified by induction from elements or deduction from a rule, etc. Endless, identity is indeterminate.

Distinguishing a ring as commutative with unity from a corecursive hyperset is the former is syntactic and the latter is semantic. A commutative ring with unity is a syntactical sequence which is analogically corecursively identifiable only as a whole in reverse sequential order. A corecursive hyperset is a semantical sequence which is analogically corecursively identifiable as a whole and as each element of the whole in reverse sequential order.

Illustrating distinction between line and ring is Gödel numbering whereby, It is desirable, however, to assign a single number to the formula [representative of a linguistic sequence] rather than a set of numbers. This can be done easily. We agree to associate with the formula the unique number that is the product of [all] the . . . primes in order of magnitude, each prime being raised to a power equal to the Gödel number of the corresponding elementary sign. The . . . formula is accordingly associated with the number.1

“Jane loves John” and “John loves Jane” have the same Gödel number, but not the same meaning, since either can be true without the other being true, love not needing be reciprocal. Gödel numbering presumes sentences are reciprocal when they need not be. “Quickly Pat ran” and “Pat ran quickly” illustrate sentences can be reciprocal, but Gödel numbering cannot distinguish which are and which are not.

The Field

Constituting a field is a recursive set, initiated by identifying a constant analogical archetype.2 Other things like it are then identified, composing a recursively identified essential set. Being such, no property is contained, tautology determining a limitless class. Ernst Zermelo and Adolf Abraham Halevi Fränkel achieve this by identifying number with location in an iterative sequence, ‘So every number except 0 has exactly one member.’3 Unambiguous as limitless, a recursive set is necessary.

Identity of elements is not a function of identity of the preceding elements. It is a function only of the initial sequential element. Thus a recursive sequence is not a function of its order, an alteration in any preceding element not altering the content of subsequent elements. Uniform, constituents are freely

substitutable with one another and relationships are freely substitutable with one another, without altering identified sequence. Same sequence in any order is constituted, any constituent acting as intensional archetype generating same intentional membership.

Distinguished is analogy and tautology. Analogy is similarity of a constituent of a set to a constituent of another set. Tautology is similarity of a constituent of a set to a constituent of the same set. Constituents of a recursive set are both intensional criterion and intentional constituent of the universal class. They are intensional criterion because any member of the class is an analogical archetype for membership in the class. They are intensional constituent because any analogical archetype for membership in the class is a member of the class.

**Conclusion**

Resolving the disjoint between quality and abstraction when substituting ‘abstraction’ for ‘consciousness’ and ‘quality’ for ‘brain,’ Colin McGinn is not mistaken speculating, ‘The reason that [abstraction] must have a hidden structure, intrinsic and essential to it, is that the unknown properties that link it to [quality] must be *internal* to [abstraction].’¹ Like physical forces, the ‘hidden structure’ of abstraction is the ability to infuse quality throughout without substituting for it. Identified recursively is the continuum and iteratively is the limit of quality. Quality can be ordered in endless ways, however, rendering any principle by which it can be ordered arbitrary, not necessary. Thus, distinguishing lines, rings, and fields are logical types. Diverse in content, lines and rings are iteratively generated, limited as such. Uniform in content, a field is recursively generated, unlimited as such.

¹McGinn. 44.
Hegel’s Analysis of Universality in the Science of Logic

Ioannis D. Trisokkas, University of Warwick, UK

S peculative logical theory, as expounded in Hegel’s *Science of Logic*¹, is that part of his systematic philosophy whose aim is to show what determinations of what there is *in truth* can be known with *absolute certainty*. It consists of three particular logics: the logic of being,* logical theory, as expounded in Hegel’s *Science of Logic*¹, is that part of his systematic philosophy whose aim is to show what determinations of what there is *in truth* can be known with *absolute certainty*. It consists of three particular logics: the logic of being, the logic of essence, and the logic of the concept. The relation holding between them is derivational or developmental in character; the starting-point of the logic of essence derives immanently from the outcome of the logic of being and the starting-point of the logic of the concept derives immanently from the outcome of the logic of essence. The specific character each logical sphere’s starting-point has is absolutely crucial for the manner in which it will develop its categories.

In this essay I wish to contribute towards the clarification of the starting-point, namely the first category, of the last logical sphere, the logic of the concept. The name that the theory employs to refer to this category is *universality* or *the universal concept* or simply *the universal*. A complete clarification of this category would require the accomplishment of three distinct tasks: (a) to explain how it emerges from the logic of essence;² (b) to provide an account of its own independent structure;³ and (c) to explain how it relates to the other categories of the logic of the concept, especially the ones that immediately follow it.⁴ The present essay will be able to fulfil only the

²WL II 217-240; see also WL 245-251.
³WL II 274-279.
⁴WL II 280-288 (particularity) and WL II 296-301 (individuality).
second of these tasks, although elements relating to the other two will be occasionally employed.

Self-identity and Utmost Determinacy

The concept appears explicitly through the result of the dialectic of causality, the last purely essentialist category in the logic of essence. Causality develops its content in three stages: (a) As formal causality, it designates the simple opposition of cause and effect; this opposition collapses because the independent existence of the effect transforms it into a simple thing (i.e. a thing which is no longer an effect) and this immediately transforms the cause into a simple thing (i.e. a thing which is no longer a cause) as well. (b) As determinate causality, it expresses a single element that distributes itself to two different substances, to the one as a cause, to the other as the effect. What determines the distribution is the character of the one of the substances; it must namely be an active substance. But as an independent substance it must become active; if this is to be achieved through stimulation, there must be another cause; obviously this leads to an infinite regress of causes and effects. (c) As conditioned causality, it designates the relation between an active and a passive substance; for it is not enough there be only an active substance for the explication of causality; the active substance acts upon a passive substance. But the active substance, as seen, exits its simple independence through its stimulation and, therefore, is itself also passive. On the other, the passive substance, as just pointed out, is absolutely crucial for there to be an active substance; in other words, passive substance is a condition of the active substance. This turns it into an active substance as well.

Causality has now become reciprocity, a structure in which a manifoldness of substances determine one another by being simultaneously both active and passive regarding one another. This mutual interaction signals the emergence of a new relation, in which the relata are bound non-causally; they emerge together in one, self-identical relation. The new relation takes the name concept and the first moment of its explicit characterization is called universality.\footnote{For a more detailed analysis of the move from the logic of essence to the logic of the concept see Houlgate (2004); in conjunction with this article read also Houlgate (2000).}

Universality denotes, initially, the simple presence of a structure shaped by reciprocity’s two dominant features:

1. The universal is absolutely self-contained or self-identical. This must be so because the relation of the substances is non-causal, they emerge together in a single unity. With respect to this relation, then, the new structure, the universal, has no other; its determination arises out of its own self.
2. But, secondly, the universal is determinate, something that requires the existence of differentiating relations and, therefore, the presence of the other. Moreover, this determinacy, as being within a self-contained element, must lack nothing. It is not the case, namely, that
there are substances which determine others and still are not determined themselves through those others. Clearly, in such realm of absolute interaction the universal must have its whole determinacy within itself.\(^1\)

As having the most determinate content, the universal is sharply distinguished from pure being, which was the only previous category that was also absolutely self-contained.\(^2\) Their difference lies in that pure being had the most indeterminate content, since its involvement in relations was explicitly denied. The analysis of that category has shown that its absolute indeterminacy is an illusion and soon becoming takes its place as the truth of being. But the vanishing of pure being results not only in the loss of absolute indeterminacy, but also in the loss of self-identity. All categories from becoming to causality are nothing else than manifestations of the immanent struggle of thought to harmonize the element of determinacy with that lost element of self-identity. The universal seems to be the exact point at which this harmonization finds its explicit fulfilment, even if only in a still undeveloped way.\(^3\)

Having specified these simplest features of the universal, the theory attempts to provide an explicit account of the manner of their unity. How exactly can it be shown that self-identity and utmost determinacy belong together in an inherent unity without any problems occurring?

**Persistence**

Universality is minimally characterized as a self-identity which has the most determinate content. The analysis of quality in the logic of being has already shown that determination is produced through negative relations among different elements. Therefore, it must be the case that, since (a) the universal is self-identical and (b) its content is constituted by negative relations among different elements, the universal persists through this spectrum of negative differentiation. Something must be differentiated in order for the ‘most determinate content’ to be produced; this something cannot, however, be the universal per se. Determinacy is not a limit to the universal, a point which demarcates its passage to an other, but rather the field in which it maintains its identity.\(^4\)

This further characterization of the universal as persistence is a feature that distinguishes it from (a) a typical categorial structure in the logic of being and (b) a typical categorial structure in the logic of essence. A category in the sphere of being has two fundamental determinations; firstly, it is immediate and, secondly, its relation to its other is one of passing over. Its immediacy denotes the fact that it initially appears as if it had a content that could be explicated independently of any reference to its opposite. Finitude, for example, develops its content without any reference to infinity. When it

---

\(^1\)WL II 275.
\(^3\)WL II 275.
\(^4\)WL II 276.
becomes evident that finitude must necessarily transform itself to infinity, this transformation does not take the form of a synthesis, but one of passing over: infinity is now what is true in being and the independence of finitude vanishes altogether. Simply, in the logic of being all being is in truth explicitly expressed each time by a single category, taken independently from its other. In contrast, the universal’s relation to the other is not one of passing over, but one of persistence. It might be the case that the content of the universal requires differentiation and manifoldness, but in the very process wherein these elements are actualized the universal explicitly persists as one and the same.¹

The characterization of the universal as persistence brings forth also its distinction from the typical categorial structure in the sphere of essence. An essentialist category also relates immediately to its other, but this relation is one of dominance, not one of persistence.² Dominance arises because relations in the sphere of essence are the result of the activity of positing: An essentialist category posits its other, to which it is irreducibly bound, so as to establish its own logical priority and superior reality over the other. The problem is that what does the positing comes into existence (or becomes meaningful) only through what is posited! Thus, the claimed logical priority and/or superior reality of the one part of a given essentialist pair over the other turns out to be only the product of an external opinion: It is simply decided, without taking into consideration the true logical behaviour of the terms, that one is prior to or superior over the other.³ In contrast, the universal, as persistence, does not relate to an other from a prior or superior standpoint. It is rather absolutely present in the determinations that are so bound in relations of otherness.

**Absolute Negativity**

The problem of the manner of the unity of self-identity and utmost determinacy has been illuminated through the notion of persistence. But the question re-emerges in a different guise: How does this persistence take place? The theory insists that the manner of the universal’s persistence in the field of its own determinacy must be explicated in terms of absolute negativity.⁴ But how can this notion be specified?

To begin with, absolute negativity is a process of double negation. The first negation constitutes the universal’s own field of determinacy and expresses a manifoldness of relations and relata. This specific manifestation of the universal is called particularity.⁵ But the universal is not just determinate, it is also self-identical and the most determinate. To achieve the explication of this aspect of the universal a second negation takes place which negates the first. What is negated is the negativity of the manifoldness present in the universal’s field of determinacy. In a later stage, speculative theory shows that this second negation brings forth a double result: firstly, the concept returns into itself from

¹WL II 276.
³WL II 276.
⁴WL II 274-275, 275, 276, 277.
⁵WL II 277, 278.
the negativity of particularity and is exemplified as an explicit totality; but, secondly, it becomes the completely other of itself, absolute determinacy or concreteness, the non-conceptual element. This structure, where the universal, through the negation of the negativity of determinacy’s manifoldness, manifests specifically these two aspects is called individuality.¹

We began this inquiry into the nature of universality by specifying that it takes over the content emerging from the collapse of essentialism and that this content is fundamentally characterized as self-identical and utmost determinate. From its self-identity we deduced its persistence in the field of its own determinacy. Next, we saw that speculative logical theory characterizes the manner in which this persistence is logically actualized as absolute negativity. This has been explicated as involving two negations. The first is expressive of a manifoldness of relations and relata; the second of the subsumption of this manifoldness under a single whole, a totality. The content of this overall movement allows the following description of the constituent moments of the concept:

(a) **Universality** is the concept’s simplest and most immediate content. It is characterized absolutely as self-identical and the most determinate. In its very purity it suppresses (although it does not annul) the specification of its determinate structure.

(b) **Particularity** specifies the determinate structure of the universal by emphasizing the manifoldness of relations and relata that are present in it.

(c) **Individuality** is the structure that clarifies that all this manifoldness belongs to a single whole. It becomes later apparent that this clarification entails both the totalization and the externalization of the concept.

Thus we reached the end of the first part of the speculative logical theory of universality. This is supplemented with a second, which focuses on the relation between the moments of universality and particularity. The question asked is this: How is it that determinacy, in being determinacy, is still a self-identical universal? The question will be discussed not from the standpoint of individuality or particularity, but from the standpoint of self-identical universality. Indeed, we are still not interested in (a) how the field of determinacy dissolves into individuality or even (b) how exactly the field of determinacy functions as such, but rather in (c) how the persisting element of universality manifests its self-identical character in the field of its own determinacy.² Simply: we want to understand what it means to say, on this primitive level, that particularity, as determinacy, is the manifestation of a self-identical universal.

¹WL II 274, 277.
The Single Movement Model: Shining Outwards

Speculative logical theory attempts to explicate the fusion of universality and particularity in the sphere of self-identical universality through the identification of the structure of universality with double shining.¹ This exemplifies the unity of two distinct operations: on the one hand, the universal shines outwards; on the other, it shines inwards. The task is to understand what these expressions mean here, how exactly their unity is achieved, and how their unity corresponds to the unity of self-identical universality and particularity. In order to fulfil this task I will first develop a model of explication that employs only the shining outwards (call it the Single Movement Model [SMM]); next, I will show that such a model is deficient; finally, it will be shown that if a second model complements the shining outwards with a second logical move, the shining inwards, the problem created by the SMM can be removed (call this second model the Double Movement Model [DMM]). My claim is that this route from the SMM to the DMM represents the exact content of the second part of the speculative logical theory of universality.

First the Single Movement Model. To begin with, the universal, as distinct from pure being, has a determinate character and hence must exhibit a structure expressive of negativity. Since all that is present in the current logical sphere is the element of universality, such structure arises out of the negative relations a certain universal develops with other universals.² This specific move entails that the universal in question also relates to a higher universal or genus. This is because (a) by relating to another universal, it becomes a relative universal or, simply, a particular which confronts an other particular, and (b) the universal should remain self-identical, namely it should not relate to an other; the positing of a higher universal is meant to save the self-identity of the universal.

The peculiarity of the SMM consists in that it understands the positing of the higher universal as further entailing that the latter, just by virtue of its higher status, cannot be described as an other to all those universals which behave to each other as particulars. For the SMM, the whole point of having a higher universal is so as the negatively related universals to be dissolved therein and the self-identity of the universal thereby be restored. As elements whose whole existence is determined by their dissolution into a higher universal, the dissolved universals are more specifically called species.

Thus the SMM holds three theses:

1. The determinacy of the universal requires that a multiplicity of universals relate negatively to each other.
2. The self-identity of the universal requires that the negatively related universals are dissolved into a higher universal.
3. The higher universal, which is the universal restored in its self-identity, does not relate to the dissolved universals in terms of otherness.

¹WL II 278-279.
The first two theses constitute the content of the *shining outwards* and represent the two fundamental features of the *single* movement of particularity as this occurs within the logical constellation of the *self-identity* of the universal. They belong to a single movement because they are both expressions of a relation in terms of *otherness*. The *shining outwards* means here nothing but relation-to-an-other. In contrast, the third thesis is not expressive of a movement toward an *other*, but rather of what results from that movement.

**The Problematic Character of the Single Movement Model**

One might argue that the SMM has provided us with a successful account of the fusion of self-identity and determinacy: The *self-identity* of the universal *in* the field of its own determinacy is achieved through an act of *dissolution* of negatively related universals. But there is a twofold problem with this model:

1. It leaves the issue of the *determinacy* of the *higher* universal unspecified.
2. It allows for a specification of this issue that would fail to give a successful account of the harmonic fusion of self-identity and determinacy in the structure of the universal.

Taken without any further clarification or addition, theses (1) and (3) of the SMM lead to an undesirable conclusion regarding the specification of the *higher* universal’s determinacy. Thesis (1) demands that the universal should relate to an *other* if it will gain determinacy; thesis (3) specifies that the higher universal is *not* related to the dissolved universals in terms of *otherness*. Thus, there are only two possible conclusions here: either (a) the higher universal is rendered *indeterminate*, which is an option that does not fit in well with the structure of universality in the present working framework, or (b) the higher universal gains determinacy through its developing relations with other *non-dissolved* universals.

The remaining second option, however, is equally ineffective, for the relations the higher universal develops with other non-dissolved universals have to be *dissolved* once more into an even higher universal.¹ This follows necessarily from the premise that relations of otherness designate the loss of the self-identity of the universal. Clearly, the need for determining the higher universal initiates in this instance a progress of universals.

This result is problematic in two ways:

1. Given the endless need for determining the ever higher universal, such progress of universals becomes *infinite*. Infinite progress, however, does not fit in well with a model that purports to explicate the *self-identity* of the universal. It implies that the latter will be *always* remaining a step *prior* to its actualization.
2. Given the endless need for determining the ever higher universal

¹WL II 279.
and the premise that the dissolved universals do not play any role in this process of determination, it seems that the progress of universals must be accompanied with the elimination of large chunks of lower-order determinacy. In other words, the higher universal will become more and more abstract. But since universality is fundamentally characterized not simply by determinacy, but rather by the utmost determinacy, a model of explication of universality which results in the elimination of large chunks of determinacy must be deemed incapable of satisfying our expectations.

In this section I have tried to show that the SMM is a deficient model for the explication of the harmonic fusion of self-identical universality and determinacy (or particularity). It is because of this failure of the SMM that the speculative logical theory of universality develops the notion of double shining as the appropriate exemplification of this fusion.

The Double Movement Model: Shining Outwards and Shining Inwards

The DMM develops a structure which incorporates the fundamental determinations of the SMM, but adds to those the shining inwards with the aim of avoiding its pitfalls. In this new model, the determination of the universal still requires the existence of simple negative relations among universals; and for the self-identity of the universal to be maintained the dissolution of the particulars to a higher universal is still needed. But instead of repeating this process with respect to the higher universal’s determination, this is achieved through a reverse movement or ‘shining inwards’, a movement that is directed back to the dissolved universals.

The shining inwards is described as the second negation; clearly, then, it acts against the shining outwards; this means it goes against both (a) the negative relationality of universals and (b) the process of dissolution of particular universals into a higher universal. But since these features of the determinate self-identical universal are not illusory, but necessary aspects of it, establishing the shining inwards cannot possibly imply their annihilation. Negating the shining outwards does not mean cancelling it. What does it mean then?

At this point the notion of character is introduced. The universal relates to the field of determinacy not only by dissolving it into a higher form of itself, but also by incorporating and permeating it as an immanently built character. This character built into the field of determinacy is the manifestation of the genus, in which the negatively related universals have already been dissolved. The manifestation of the character of the genus in its own determinacy constitutes the exact content of that which speculative logical theory calls the shining inwards of the universal.

---

1WL II 279.
2WL II 278.
One way to illuminate this further is to clarify how the relation between ‘shining inwards’ and ‘shining outwards’ should not be conceived. Hegel is not saying that the manifestation of the character of the genus is the result of the process of dissolution of species into a genus. He does not namely argue that the shining outwards and the process of dissolution produce the shining inwards. Nor is he saying that the genus causes the process of dissolution to manifest its character. This kind of relation has been left behind for good with the demise of essentialism.

What Hegel does say is that for the determinacy of the self-identical universal to be successfully explicated there should be not only a process of dissolution of species into a genus, but also the manifestation of the character of the genus into and through the process of dissolution. Neither of those processes cause the other; in fact, they represent one and the same process. Nonetheless, they must be characterized in this twofold manner, because the simple notion of a process of dissolution of species into a genus does not fulfil the present task of speculative logical theory, that is, the explication of the determinacy of the self-identical universal.

Thus, the shining inwards is the negation of the shining outwards, not in the sense that it cancels it altogether, but rather in the sense that it cancels its self-subsistence. The theory purports to have shown that to think of a determinate self-identical universal means (a) to think of a process of dissolution of particular universals into a higher universal and (b) to think this very same process as the manifestation of the character of that very same higher universal. The structure of double shining, therefore, is constituted by two acts, which are (a) both necessary and (b) complementary of each other.

This double structure is necessary because the process of dissolution, taken in isolation, does not suffice for the explication of the self-identity of the determinate universal. But why does the process of manifestation of character not suffice for such an explication? Why must it be complemented with a process of dissolution? The answer is this: The manifestation of the character of the universal is made through determinacy and particularity, which require the existence of negative relations among distinct universals; but the whole point of the theory of universality developing a second part has been to show that despite this differentiation and manifoldness, the universal can still be characterized as self-identical. Something extra, therefore, must be said in order to move beyond such differentiation and manifoldness. To say that the distinct universals are manifestations of the character of their single genus is exactly to say they are dissolved into a higher universal. Having two distinct universals standing alongside one another does not entail that they express the self-identity of the universal; one has to mention this higher universal, and, in particular, to mention it as something that has a distinct existence from either one of the particular universals. The complication stems from the fact that the distinctness of this third universal should not be taken as requiring the elimination of the first two universals. But there is only one way for this to be avoided, that is, by specifying that the third universal manifests its character in the first two universals and that, in fact, it is nothing else than the process of that manifestation.
The Absolutization of the Higher Universal

Let me now briefly refer to an interesting and controversial ramification of the preceding discussion. The universal has been treated as the locus where the particular universals dissolve and as the character manifested in them and their interrelations. One might argue, though, that conceiving the higher universal in this way does not exclude the possibility of there being a multiplicity of higher universals, totally unconnected with each other and still totally determined. Indeed, given that its own character and species determine it absolutely, there would be no need for a certain higher universal to relate to other higher universals. In other words, this higher universal’s complex relation to its own species suffices for explicating its self-identity and determinacy.

This picture, however, entails that the universal would now lose its self-identity and self-containment. This is because the existence of a multiplicity of universals, which are totally unconnected with each other, implies that the universal is now disparate and irreducibly divided. Yet, if you recall, it is a fundamental premise of the current problematic that the universal is self-identical and self-contained. The fact, then, that a higher universal might be absolutely determined through its own species does not suffice for regarding it as not relating to other higher universals, that is, if it is posited that other higher universals do really exist.

Had we assumed, therefore, that at a given moment there exists a multiplicity of higher universals, speculative logical theory must explain how the universal’s self-identity arises out of it. There are two options available: it either (1) employs once more the notion of dissolution into an even higher universal or (2) affirms that there remain no more universals which are not dissolved into a certain higher universal. The first option, taken in isolation (to wit, as it operates in the SMM), must be rejected, for – as already shown – it achieves self-identity without achieving determinacy, and vice versa; but even if it is taken in its more advanced state, that is, as it is exemplified in the DMM, it would still be problematic because it would still not possess the power to exclude the postulation of the above assumption. The latter can be excluded only if the DMM is, in its turn, complemented with the second option above. Thus, if the theory is to achieve its programmatic aim, there must be a point in the dialectic where the process of dissolution stops and all universals are shown to dissolve into a certain higher universal. Next, this higher universal will obtain its determinacy in the manner described in the DMM. Just because there can be no universals that are not dissolved therein, this latter universal should now more succinctly be called the highest universal.

Thus, the logic of Hegel’s argument demands the absolutization of the higher universal, its transformation into the highest universal. This new element does not change much of the essence of the DMM, except that it now becomes crystal clear that the shining inwards, which differentiates this model from the SMM, takes place from the standpoint of the highest universal (and not from the standpoint of any higher universal whatsoever). The character which the particular universals manifest is the character of their own genus; but the latter is now revealed to be the highest universal into which they dissolve themselves.
Conclusion

Speculative logical theory is that theory which aims at the presuppositionless discovery of the kind of thought that could provide human beings with knowledge of what there is in truth. The collapse of essentialism brings forth conceptual thought as the object of analysis. Universality is the structure that conceptual thought exhibits when it first appears. Its fundamental characterization as self-identical and fully determinate gives rise to a certain tension, for while determinacy requires the manifestation of otherness in order to be explicated, self-identity requires the suppression of otherness. This tension, however, has a completely different character from the one it had in essentialist thought. There, thought understood reality as being constituted by relations of dominance and externality. In the sphere of the concept, in contrast, such relations play absolutely no hegemonic role. This difference in character on the plane of relationality distinguishes the tension found within each sphere. While in the sphere of essence thought is driven by a primordial division between self-identity and determinacy, in the sphere of the concept what shapes the whole dialectical development is the primordial unity of self-identity and determinacy. The outcome of the logic of essence was that this primordial division cannot be sustained, for the related terms collapse into one another and, therefore, each fails to establish a priority over the other. The logic of the concept begins with this exact picture, namely that the related terms coexist in perfect unity, in a relation of love. The tension, therefore, that we encounter in the sphere of the concept is, as it were, a ‘friendly’ one. Speculative theory does not proceed by looking for a way to resolve the tension between self-identity and determinacy, because this has already been resolved in the outcome of the logic of essence; it only aspires to understand the (already existing) harmony between these two fundamental determinations.

The first step toward this understanding has been made in the sphere of self-identical universality and the unity of the determinations has taken up the structure of double shining. The shining outwards exemplifies the universal’s relation to an other and has two sides: it is (a) a manifoldness of particular universals that relate negatively to one another and (b) a process of dissolution of this manifoldness to a higher universal. The shining inwards exemplifies the highest universal’s relation to the negatively related, dissolved universals; this is not a relation to an other, but a manifestation of a character, the character of the genus, through the manifoldness of particulars. One could also put it thus: The shining outwards exemplifies a relation of differentiation among particular universals and subsumption to a higher universal; the shining inwards exemplifies an event of inherence of the highest universal in the particular. The relation between the two sides of double shining is one of strict implication: It is logically impossible to have the one without the other.

Bibliography

An Anthology of Philosophical Studies Volume 2


Part V

Ethics and Political Philosophy
ew insights from the sciences, especially from the life sciences, provide new ways of understanding, controlling and manipulating human functioning. Elementary characteristics and abilities, both physical and mental, can be regulated more easily and more specifically than ever before. These possibilities raise questions about “human nature” and draw particular attention to the basic and crucial question of philosophical anthropology, “What is humanity?”. However these questions—What is humanity? What is human nature?—are urgently in need of conceptual clarifications. A quick look at the prevalent debates shows that the use of core concepts in the debates—such as enhancement or nature—is profoundly confusing. To avoid the negative, sometimes disastrous repercussions of unclear concepts in these discussions, some conceptual ground-work is necessary. In discussing these fundamental concepts I hope to improve the solidity of contemporary debates about the urgent ethical as well as anthropological problems concerning human enhancement.

This philosophical work refers to a concrete bioethical question. But bioethicists¹ need philosophical clarification of their reflections about concrete problems. Otherwise their considerations and proposals risk remaining superficial or—at the very best—purely pragmatic. Hence philosophy contributes to a particular science as a meta-science insofar as it facilitates an understanding of what is really at stake in the debates about human enhancement and where exactly the hidden core-problems are to be found.

¹There is an argument to be made that bioethicists belong more to applied sciences than to philosophy, but I do not pursue it here.
The Context

There are two main fields in which impressive new insights from the sciences give rise to new options of human intervention: genetics and the neurosciences. In both areas new medical competences not only offer new means of therapy but also allow for multiple uses in healthy people. This use of medical treatments in the absence of any indication of relevant symptoms is often called “enhancement”, because the aim of these interventions is not to restore health but to improve or to enhance certain human characteristics or capacities above a normal level. This already reveals a problematic term, namely “normal level”, which will be discussed later. This possible use of medical interventions outside of the usual area of healing diseases brings with it several difficult and problematic consequences. Two brief examples may help to illustrate this thesis.

**Genetic engineering:** The growing insights into the functioning of genomic processes (especially in reproduction, but also in the development of certain traits), in combination with new means of voluntarily influencing them, allow for interventions that go far beyond any known until now. Stem cell intervention to alter the expression of certain genes or to modify the human genome (of an individual, but, through reproduction, of the offspring as well) might even be used to introduce new genes—or genes from other species—into the human genome. One aim could be to produce resistance to diseases such as AIDS or other viruses, but alterations in physical appearance or performance are also becoming feasible.

**Neuro-enhancement:** Another field of new interventions is that of medications that influence human brain functioning. There are many psychotropic substances; here I consider only three. Ritalin™ is usually administered to children suffering from ADHD (attention deficit hyperactivity disorder) to improve their ability to concentrate. But healthy people might also be attracted to the idea of being able to concentrate even better. Another drug prone to multiple uses is Modafinil™, usually administered to narcoleptics to help them stay awake. And, as a third example, we can take Prozac™, an antidepressant (a so-called selective serotonin reuptake inhibitor). But in healthy people as well this medication can cause a shift in the spectrum of the possible mood states of an individual. This enables a change from any given state $A$ to a better state or mood $A^+$. The consequences of this mood-brightening are manifold and might ultimately lead to concrete results such as more success in the job and a better fiscal compensation.

---

1A comprehensive survey of these interventions can be found in Merkel et al. (2007).
2Cf. President’s Council (2002).
3The consequences of the use of Prozac have been intensively discussed since the successful 1993 book *Listening to Prozac* by the American psychiatrist Peter D. Kramer (Kramer, 1997). A few years ago a new book with the title *Better than Prozac* by Samuel Barondes
Even though these fields of genetic engineering and neuro-enhancement differ in many respects—different degrees of reversability, different degrees of invasiveness—they should be discussed together, as they both give rise to the same ethical problem: *How far should humans intentionally intervene in the physical underpinnings of their existence?*

**Consequences of these Examples**

These medications or interventions, which may alter people’s physical functioning to achieve mental or physical states formerly out of reach to them, can be seen as ways of *deeply altering human beings* — whereby obvious changes are realised in individuals in both the physical and the mental sphere following the administration of a certain drug or gene therapy. Here indeed one finds a massive human intervention in the material basis of a person’s physical and mental existence. Combined with a conviction that most modern people hold dear and which may be called "biological materialism"¹, these changes often are seen as threatening important characteristics of human existence. I use the term 'biological materialism' to designate the widespread believe that the biological sphere—our physical brain and our genes—is to a very high degree responsible for what we are. The complementary sphere—the influence of education and culture on personal development—is often rather neglected compared to the biological sphere.² If the biological sphere underlying our individual and human identity becomes increasingly better understood and hence more changeable, many people feel deeply unsettled and see this development as a growing menace.

**Some too Narrow Considerations of the Problem**

Contemporary bioethics very often asks the wrong questions. As interesting and as useful it may be to think about the just distribution of and equal access to scarce resources³; or to discuss the risks and safety problems of these new means⁴; as important as it may be to discuss questions of fairness and cost⁵— the real and currently pressing problems of the enhancement-debate lie elsewhere, as there is much more at stake than how to regulate the access to these new means. Habermas⁶ defines the core of the problem of ‘liberal

---

¹This term is used by Kramer (1997).
²Hence articles with headlines like ‘Gene for homosexuality discovered’ are seen much more frequently than claims of moncausal explanations for certain individual characteristics from the cultural sphere.
³Buchanan et al. (2001).
⁴E.g. Fukuyama (2002).
⁵For a critical approach see Savulescu (2006).
⁶Habermas (2001). All translations from German by JC Heilinger.
eugenics' as a possible ‘self-transformation of the human species’ (42) and thus identifies a problem that is relevant not only for our specific cultures but also for humanity in general. This concerns

... our self-understanding as belonging to a certain species. It is not about culture, which is different everywhere, but about the conception that different cultures have ‘of man’ who—in anthropological universality—is everywhere the same. (72)

These examples and others of interventions aimed at enhancing the human species show (according to Habermas) that this ‘mechanisation (Technisierung) of human nature gives rise to a modified self-understanding regarding the ethics of the species (gattungsethisches Selbstverständnis)’ (76). Indeed, Habermas expresses a serious concern with the probable outcomes of the use of these new possibilities, which might cause a wide range of changes. Nevertheless, Habermas’ doubt still remains rather unspecific when he talks about the ‘mechanisation of human nature’, or a ‘modified self-understanding regarding the ethics of the species’. The concerns also are not concrete enough when people refer to “enhancement” in general as a somewhat spooky practice of eugenic or Frankenstein-like intervention into human organisms. In order to secure a substantial foundation for adequately addressing the possible argumentative reasons for or against these specific interventions, one has to consider two problematic concepts underlying these concerns: the concept of “enhancement” as well as the concept of “(human) nature”.

Enhancement

The entire philosophical and bioethical debate is labelled the “enhancement-debate”, but what exactly is an enhancement? Every definition of enhancement necessarily remains unclear as there is no distinct natural kind of “enhancing actions”. Enhancement can be (1) understood in contrast to therapeutic interventions. But this distinction presupposes a clear boundary between health and disease, which is necessary to support a clear opposition between therapy and enhancement. Is there such a clear distinction between

1Habermas focuses on genetic interventions, which can be seen as the “deepest” possible intervention into the human species until now since the consequences of these interventions can be passed on to future generations. But Habermas’ ideas are nevertheless valid for less far-reaching interventions as well, but to a lesser degree. A lucid discussion of the problems arising with liberal eugenics can be found in Agar (2004).
2The underpinnings for Habermas’ concern lie in an asymmetrical communicational situation between modified and non-modified human beings. I will not further discuss this particular argument here, as it seems to me to only be valid under strong presuppositions that Habermas makes in the context of Transzendentalpragmatik.
3This is the approach of the President’s Council of Bioethics (2003). But there are morally suggestive implications built into this distinction from the very beginning, as therapeutic interventions are morally in no way problematic while enhancing interventions then appear to be prima facie morally questionable.
disease and health? Certainly not. Another way of understanding enhancement is (2) in terms of the normalcy related to human species. But does so-called “human nature” provide a sufficiently clear basis for defining which human actions alter human nature? I argue that human nature as such is not capable of providing a sufficient basis for defining enhancement, because it is a completely unclear concept, prone to many abuses by all sorts of argumentative tricks.

For these reasons, I propose—somewhat provocatively—defining a human action an enhancement if and only if it is an intentional intervention in the human body or mind which is (subjectively) positively evaluated. This “minimal definition” of enhancement has several consequences, one of which is usually seen as a disadvantage: it comprises nearly all human actions. Even such actions as having a cup of coffee with a decent breakfast, or learning, can fall under this definition. But I consider this alleged disadvantage as the real significant advantage of the so-called minimal definition, because it allows us to group together all human actions taken on purpose to improve our functioning. Prima facie there is nothing morally questionable about doing something to improve ourselves. These attempts rather seem to be typical and specifically human kinds of actions. Of course further specifications must be made, as not every enhancing action is equal to every other. But—and this is the real import of the proposed minimal definition—the fact that an action is an enhancement does not by itself say anything about its moral qualities. The main problem of the term enhancement—that it mostly has the connotation of something morally questionable—can be eliminated by a purely descriptive term which allows for a sober discussion of the problems at stake.

After arguing for the thesis that enhancement is not necessarily bad or wrong one can turn to the supposed core problem of the enhancement-debate. It is feared, as Habermas wrote, that certain enhancements might change human nature. In light of this, certain people conclude that we should not use these enhancements. What does this mean?

The “Argument from Human Nature” against (Human) Enhancement

The “argument from human nature” against certain human enhancements, seen as a problematic intervention into human existence, is often summarized in one of the following ways.

---

1I tried to show this in more detail elsewhere, cf. Heilinger (forthcoming). The concepts of health and disease allow for multiple and heterogeneous definitions, so that a supervening distinction between therapy and enhancement necessarily inherits all of the difficulties of the first distinction.
3This will be discussed below.
4Cf. Heilinger (forthcoming).
5Besides Habermas (2001), one can find arguments based on a normative understanding of (human) nature not only in religious contexts, but also e.g. in Fukuyama (2002), President’s Council (2003), Siep (2004), and Kass (1997).
(1) We should not enhance our bodies and minds above a certain degree given or allowed by human nature.

or:

(1') Nature should be the guide as to which interventions in our bodies and minds are legitimate and which are not.

or:

(1'') Out of respect for human nature we should refrain from intervening in our natural equipment.

But none of these formulations of the argument says anything helpful. As long as we do not clarify what we mean by (human) nature, we cannot rely on it to give us guidance in decision-making processes. A deeper insight into the possible meanings of the term “nature” is necessary. But this begs the question. Why should we care at all about the concept of nature? What are the advantages of the term nature thought to be, why is it used so frequently in ethical debates? “Nature” is a highly attractive term because it seems to relate to a universal, objective, eternal and egalitarian entity. Hence nature would indeed be an interesting candidate for the foundation of morality. A morality founded in this way would be regarded as trans-culturally valid, because, in general, nearly nobody could object to the thesis that nature is something fundamental, pertaining to all (human) beings, no matter who or what they are. But there are (at least) four serious problems with the term nature:

1. The term is ambiguous.
2. It is multi-dimensional.
3. It exists in different degrees and the differences are not clear cut.
4. It is generally positively connotated without sufficient reason.

(1) The term “nature” is ambiguous, which becomes quite plain as soon as one tries to determine its opposite. Among those opposites of nature or the natural with moral connotations are such heterogeneous concepts as the “unnatural”, the “degenerate”, even the “perverted”. As far as these moral connotations relate to the normal one can name as opposites the “artificial”, the “abnormal”, and the “pathological”. On an understanding of “natural” as something close to authentic, its opposites would comprise the “untrue”, “affected”, or “mannered”. If “natural” stands for something close to harmonic or proportional, its contraries are the “deformed”, the “inorganic”, and the “dissonant”. These examples

---

1In the tradition of analytic philosophy I will not talk about nature but rather discuss the use of the term “nature”. Historical reflections both about nature and the use of the term “nature” are skipped here, although they provide helpful insights. The following argumentation nevertheless owes insights to John Stuart Mill’s essay Nature (postum 1874) and the thoughtful analyses of Dieter Birnbacher (2006) on the term “naturalness”. A critical contemporary approach to human nature is also discussed in a collection entitled “Is human nature obsolete?” (Baillie/Casey, 2005).

2For this cf. Birnbacher (2006, 42sqq.).
are an impressive demonstration of just how ambiguous the term “nature” is.

2 The examples also indicate that “nature“ plays a role in highly different fields and contexts of argumentation, where it has more than one “dimension”. Upon closer examination we can distinguish two dimensions: a genetic dimension relating to the emergence or formation of a certain being or trait, and a qualitative one relating to the character or composition of something. ¹ Obviously both spheres are related, but they have different outcomes. Here I simply want to point out how great a difference it makes to look at something from a genetic or a qualitative point of view.

3 The third problem is that nature is not a distinct, clear-cut predicate which can be clearly attributed to something. Instead there are different “degrees” of naturalness. In connection with the second problem described above, we will see these different degrees of naturalness in both dimensions. A thing could have emerged more or less naturally (for instance organic food grown under highly artificial circumstances compared to ordinary farming) or can be qualitatively more or less natural (for instance organic food grown under highly artificial circumstances compared to “industrial”, artificially flavoured food).

4 The three problems named above are part of another problem, which is the unsubstantiated positive connotation of the term “natural”. At least in the context of Western industrialised cultures natural goods are often preferred to artificial items. It is one of the basic and widespread convictions in our society that the natural is good. ² One can see this in the growing field of organic food stores, in the organisations for environmental protection and so on. People often try to orientate themselves with reference to what they assume to be natural. But—as positive and important as the intentions, aims, and outcomes of this position may be—it nevertheless remains unclear what precisely is natural, what is nature.

What is meant in referring to “Nature”?  

Those voices in the debates about human enhancement that use the term nature want to express something specific. The desire to contribute to the debates arises from a strongly felt intuition. To either adequately criticise these intuitions or to adequately strengthen the arguments underlying them, one has to try to understand as clearly as possible what is hiding behind the term

¹For the genetic dimension cf. the crystal-clear explanations of Birnbacher (2006) p. 9–13, for the qualitative dimension p. 13–16.
²That this positive connotation of nature is contingent is shown by a brief survey beyond out here and now.
“nature”. With the help of the principle of charity one can try to clarify the role of this term in the “arguments from nature”.

Most of the people using arguments from nature in the debates about human enhancement do so in support of a rejection of these interventions. That gene-therapy, neuro-enhancement or brain-machine-interfaces are considered to be “against nature” thus expresses a basic intuitive concern that is not further specified. Hence it is an insufficiently founded attempt to provide orientation suited to the implicit convictions that a person values. Nietzsche reproaches those providing arguments from nature:

> Indem ihr entzückt den Kanon eures Gesetzes aus der Natur zu lesen vorgebt, wollt ihr etwas Umgekehrtes, ihr wunderlichen Schauspieler und Selbst-Betrüger! Euer Stolz will der Natur, sogar der Natur, eure Moral, euer Ideal vorschreiben und einverleiben.“ (Nietzsche, 1886, 22)

Most of the time “listening to nature” tends to result in establishing a borderline that is not to be crossed, because doing so would deeply confuse the fundamental convictions intuitively felt and emphasised. One simple explanation of why contemporary Western people often tend to link their basic intuitive convictions to nature is that most people experience positive impressions when, for example, spending a Sunday in the woods. This experience many Western people appreciate but—under the contemporary conditions in our modern societies—often lack. Nature is in general seen as something good which has not (or not yet completely) been destroyed by culture and the negative influence of humans. The personal wish to escape from the technological world instead of enforcing it is the reason behind the formulation of arguments “from nature” against enhancements. Seen from the perspective of Kulturkritik, the term nature is used here as a substitute for a better world, idealised in contrast to the despised reality.

In identifying this basis for arguments from nature I do not want to oversimplify these attempts, nor do I want to make them ridiculous. There may be good evolutionary reasons to connect positive normative valuations to experiences that give rise to comfortable or pleasurable feelings (instead of preferring those that raise the stress-level and suffering in an organism). In these contexts nature is seen as something that generally supports life and

---

1This is not always true. Even arguments from nature or from religious contexts can promote enhancing interventions, cf. Peters (1996). Cf. also Bostrom & Sandberg (forthcoming). Nevertheless most of the time arguments from nature are brought up by so called “bio-conservatives” such as the Leon Kass and Francis Fukuyama.

2"While you pretend enchantedly to read the canon of your law in Nature, you want something quite the contrary, you queer stage-players and self-deluders! Your pride wishes to dictate your morals and ideals to Nature, to Nature herself, and to incorporate them therein."

3The contemporary modern technical Lebenswelt has been painted in dark colors impressively by Horkheimer & Adorno (1947) und Jonas (1984).

4The attempts to base ethics on an evolutionary foundation—following Thomas H. Huxley—can be understood as supporting this claim. Furthermore in his latest book Ernst Tugendhat discusses the evolutionary advantages of a biological foundation of fear in the essay “Unsere Angst vor dem Tod”. Tugendhat (2007), pp. 159–175.
causes positive feelings, and \textit{this consequence} might give orientation in moral questions. In this way it is not \textit{nature} as such that is responsible for these consequences, but just one particular part of something that could be called nature. In other words: reasons to keep things as they are “in nature” only follow from a positive, idealistic conception of nature.\footnote{A simply positive understanding of nature is inadequate, as John Stuart Mill has shown with impressive clarity (cf. Mill, 1874, 31, 33 et passim).} In other contexts, even those claiming a positive understanding of nature would have to concede that sometimes acting \textit{against nature} might be good (for instance in fighting against natural events as tsunamis, plagues or “natural” human actions such as rape). Here the distinction between the spheres of descriptive and the normative becomes obvious.

\section*{Conclusion}

I have argued that speaking of nature in contexts of moral orientation in general and human enhancement in particular is unhelpful and sometimes misleading. I cited some of the crucial problems with the term “nature” and claimed that the use of this term always calls for further clarification. Thus the term can and should be replaced by other, more precise expressions that make its content explicit. It would be of benefit for these debates if the term “nature” were to no longer figure in arguments about moral problems, as it causes considerably more confusion than helpful argument.

By contrast, the problems with the concept of nature listed here allow for a better understanding of this term in the debates about human enhancement. I have argued that a simple reference to the concept of human nature cannot be considered as a rational or reasonable argument, but nevertheless—with the help of the distinctions made—conceptual clarifications can be developed. When the term is used either for and against modifications of human nature, the following set of questions should be considered.

\begin{itemize}
  \item Are you referring to the genetic or qualitative aspect of what you call nature?
  \item Do you mean nature as opposed to the artificial, the degenerate, the affected, or the deformed?
  \item Do you think nature is \textit{per se} something good, worth preserving? Is there not something in nature which could be regarded as worth overcoming?
  \item Are there other ways of expressing your concerns about the intervention, avoiding reference to nature?
\end{itemize}

With these questions I have not solved the moral dilemmas and problems around human enhancement. Nor have I stated what human nature is. But I have shown how the problems and ambiguities of the notoriously difficult term can turned to advantages once they are precisely identified. It is progress to
properly identify the problematic aspects underlying an ethical debate. Hence the ambiguities listed should be kept in mind in any discussion that refers to human nature. They will help to clarify the substructure of the arguments in the enhancement debates and in doing so allow a proper assessment of their argumentative strength.

Particularly as it pertains to the discussions about human enhancement, one should not hope to gain moral orientation from human nature. In addition to discussing the question “What is humanity?”, it becomes urgent to address the broader question “What do we want humanity to be?”

**Bibliography**


Should we believe our Ethical Theories?  
Virtue Ethics and Self-effacement

Joel A. Martinez, *Lewis & Clark College, USA*

Many philosophers are attracted to modern ethical theories because they claim to offer practical guidance concerning what to do in particular situations. It is interesting and puzzling to some, then, why proponents of virtue ethics have criticized this very aspect of modern ethical theories. Virtue ethicists argue that modern ethical theories aim to give direct guidance about particular situations at the cost of offering artificial or narrow accounts of ethics. In contrast, virtue ethical theories will guide action indirectly by helping develop one’s understanding of the virtues – but, the theory will not provide answers as to what to do in particular instances. We have to figure that out for ourselves. This raises the question: Is the virtue ethical account of action-guidance defensible?

In this paper, I explain one initially powerful objection against the virtue ethical account of action-guidance and respond to it. Objection is that virtue ethics is self-effacing, or in some way recommends its own rejection. Only recently has self-effacement been seen as the basis for an objection to virtue ethics (Keller, 2007; Hurka, 2003). It is worth investigating whether there is anything to the objection and how a virtue ethicist should respond.

A larger worry looms for virtue ethicists. If virtue ethics is self-effacing, then proponents cannot avail themselves of one compelling argument in favor of the turn towards virtue ethics. That argument runs roughly as follows:

\[
P_1: \text{If a moral theory is self-effacing, then insofar as it is the theory is regrettable (i.e. all things being equal we should prefer theories that are not self-effacing).}
\]

\[
P_2: \text{Consequentialist and deontological theories are self-effacing.}
\]

\[
C_1: \text{So, consequentialist and deontological theories leave us with something to regret.}
\]
P3: Virtue ethical theories are not self-effacing.
C2: So, insofar as virtue ethical theories are not self-effacing we should pursue virtue ethical theories over the other two options.

If the objection that virtue ethics is self-effacing holds, then P3 is false. In this paper I argue that a correct understanding of the relationship between a theory of virtue and action guidance reveals that P3 is true.¹

I first explain what it means for a theory to be self-effacing. I distinguish between modestly self-effacing theories, immodestly self-effacing theories and theories that recommend indirect guidance. I explain why many philosophers think there is something wrong with a self-effacing theory and why this is commonly seen to be a problem for deontological and consequentialist theories, but not virtue ethics. Next, I explain why some philosophers think P3 is false. I claim that a fairly simplistic account of action-guidance underwrites this charge. Finally, I respond to the objection that virtue ethics is self-effacing. I argue that virtue ethics is not self-effacing, but only indirectly action-guiding.

One important goal of a moral theory is to offer a criterion of right action. A criterion of right action tells us what makes an action right. Taking the account of right action as the heart of a moral theory, we can consider what it means for a theory to be self-effacing. In the strongest sense, to say theory X is self-effacing is to say that by X’s own lights if one believes the criterion of right action X offers, one cannot perform the right action. In a weaker sense, saying X is self-effacing is to say that in many cases if one thinks about what one is doing under the terms of X, one will not be able to perform the right action. In either sense, X is self-effacing because, as a result of being convinced by X, one sees that one should act from some motivation other than the considerations that make an action right. Both senses need some explanation.

Most of us know that occasionally we have to compartmentalize our thinking.

Consider Bob the hedonist. Bob holds the view that what’s best is to do what would give one the most pleasure. For Bob, listening to music and playing tennis give him great pleasure. However, in order to get pleasure from these activities Bob cannot focus on the pleasure. Bob must focus on things like the melody or the quality of his swing. While listening to a song, if Bob considers the pleasure he is getting, he will be distracted from the pursuit of pleasure. He has now missed part of the song and is alienated from the activity in a way that precludes him from getting pleasure.

It is important to distinguish between different ways we might compartmentalize. We might compartmentalize the way Bob does. Bob is indirectly guided by his aim of getting pleasure. He does not need to hide it from anyone, he just cannot always think about it when he’s pursuing it. Though indirection is required, there is no conflict between the thoughts Bob’s

¹Note that I’m not giving an argument that P2 is true. I’m not sure it isn’t, but that’s another issue. I would be happy if P2 turned out false because we could then get on with the job of connecting moral theories more directly to education and action guidance. In any case, P2 is not required for my defense of v-ethics. This is because all sides in the debate hold that P2 is true.
hedonism requires him to have (thoughts like ‘Remember to follow through’ or ‘Notice how the theme changes slightly throughout the song’, etc.) and Bob’s hedonistic theory. I call theories that have the implication that mere indirection is sometimes required indirect theories.

There are other types of compartmentalization that raise more interesting questions. The other types typically arise from considering how deontological theories and consequentialist theories guide action. Put simply, deontological theories explain that an action is right if and only if the action conforms to a correct moral rule, while consequentialist theories explain that an action is right if and only if the action brings about the best available state of affairs.

The need for compartmentalization arises when we recognize that in many cases one ought to be motivated directly by a concern for another person (e.g. in cases of friendship). In these cases, if one is acting on the impersonal motive of conforming to a rule or promoting valuable states of affairs, then this does more than simply distract one from realizing the value in question. The motive the theory provides conflicts with the sort of motive required to realize the good in question. For example, to promote good consequences, I often need to be motivated by concern for my friend, thought of as a unique and special person, not just as another consequence. Here, what keeps me from thinking about the theory is not simply the concern to stay focused on the situation. What keeps me from thinking about the theory is the need to manage the potential conflict between my reasons for acting and the motive my reason tells me to adopt in particular cases. I call theories that require this sort of compartmentalization self-effacing theories.

There are more and less extreme ways for a theory to be self-effacing. On the most extreme version, the theory recommends its own rejection. This has typically been thought to be a consequence for certain versions of consequentialism, most notably act-consequentialism. The most commonly cited reason is that in only a very limited number of cases (e.g. a president deciding whether to send in more troops) is consequentialist thinking allowed. The result is that it follows from act-consequentialism that the vast majority of people should never think in terms of act-consequentialism. If this is right, then according to the act-consequentialist individuals should learn some theory other than act-consequentialism -- deontology virtue ethics or, maybe, rule-utilitarianism. Of course, which theory we teach will depend on the expected payoff. If a theory recommends its own rejection in this extreme way, a theory is immodestly self-effacing.

A less extreme way for a theory to be self-effacing is for the theory to have the implication that in a large number of cases one should be motivated by considerations other than the ones that, according to the theory, make the action right. One might hold that there are many cases where it is unproblematic to be motivated directly by the considerations that make the act right. At the same time, however, there can be certain kinds of cases where one must insulate one’s motives from one’s view of right action. For example, according to a version of Kantianism, to be motivated by the thought ‘Lying is wrong’ or ‘Lying is prohibited’ is permissible. However, in other cases, for example when visiting a friend in the hospital, the theory might imply that one
must insulate one’s reason for visiting (‘Friendship requires that I visit my friend’) from one’s motive (‘I visited because I care about Tom’). If a theory recommends that we sometimes act from considerations other than those that make the action right, then the theory is modestly self-effacing.

We can now see that a theory might require indirection, complete rejection (immodest self-effacement), or partial rejection (modest self-effacement). So far, a theories’ being self-effacing looks to be a quirky aspect of human life. But, what turns these reflections into an objection to a theory? Philosophers find self-effacement regrettable for two primary reasons. The first applies to immodestly self-effacing theories and the second to the modest versions.

The first reason concerns the practical aims of moral theory. The idea is that moral theories are supposed to help us figure out how to live. If a theory is immodestly self-effacing, it does not do that. It recommends its own rejection, so it cannot help me think about how to live. The obvious response is to say that even a self-effacing theory helps me figure out how to live, it just has the odd result that the theory guides me to a life that involves rejecting the theory. This response. To see why consider the objection again.

The objection relies on two related ideas. The first is that in order for a moral theory to say something about how I ought to live it must get into some of the details of the kind of life I am living. Consider an analogy with parenting. Imagine growing up in the following way. Every time you ask your father for advice he says ‘Go ask your mother.’ Imagine, further, that he’s right -- Mom always gets it right. If you had followed Dad, your life would likely be a mess. Of course, in some sense you did follow Dad, but we’d hardly say that Dad helped you grow into the person you are. Dad didn’t raise you. Dad didn’t help you figure out how to live. Dad wasn’t really there at all. In this same sense, an immodestly self-effacing theory is not really helping you figure out how to live.

One could respond that when a theory is immodestly self-effacing there is justification for it. The theory contains the normative justification that explains why one should adopt the particular motives and viewpoints one should, even where they are ones that preclude one from being motivated by the theory. On the parenting analogy, its as though Dad did not simply ignore you, but instead explained his reasons for this odd practice.

This raises the second related idea. Part of the problem with accepting self-effacement is that to do so we must overlook the fact that being moral is an ongoing project taking place over the course of a life. To expect a completed moral theory that can explain why one ought to adopt such-and-such general viewpoint or character and then disappear from one’s mind is to expect something undesirable. I take it that this is what Michael Stocker (2002) had in mind when he asked ‘How (if the theory is self-effacing) will one be able to check up on oneself to see how one is doing?’ Of course, we wouldn’t think one needed to check up on oneself if we took seriously the idea that one can start out with a completed theory from the beginning and apply it correctly. The desire to avoid immodestly self-effacing theories shows that, if pushed, few ethical theorists take the idea of ‘the completed moral theory’ seriously anymore.
The second reason philosophers find self-effacement undesirable focuses on the modest versions of self-effacement. In cases of modest self-effacement one runs the risk of psychological conflict, which many find morally problematic. Where one must insulate one’s reasons from one’s motives, one cannot be moved by the things one values and one cannot value the things by which one is moved. In short, modest self-effacement threatens our psychological unity and makes it hard (or maybe impossible) for us to live a harmonious life. We should ask ourselves not whether it is possible to live with such a conflict, but whether we can hold that up as any kind of moral ideal. On reflection, I think many of us will hold that we cannot.

The above considerations give us compelling reasons to look for theories that are not self-effacing. We have already seen why it is standard to think that consequentialist theories and deontological theories are, at least, modestly self-effacing. In the case of close personal relationships, we cannot be moved by rules or the production of the best consequences. Virtue ethical theories have provided an attractive third option because they seem to offer a way around this problem. Consider the basic virtue ethical account of right action:

R: An action is right iff a fully virtuous agent would characteristically (i.e. acting in character) perform it in the circumstances (Hursthouse, 2000)

This account of right action appears to avoid the problem of self-effacement because it expresses a view about what the virtuous person would do: the virtuous person would visit the hospital from a genuine concern for her friend. The virtue ethics account of right action explains right action in terms of motives and that makes it impossible for the virtue ethical account of right action to come apart from the virtue ethical account of good motives. In the next section we will see why some philosophers think this is not the case.

It is worth taking a second look at the virtue ethical account of right action. Simon Keller (2007) has argued that if virtue ethical theories are to avoid self-effacement, it must be because such theories are compatible with the following,

A: Its always acceptable to be motivated by the considerations that, according to virtue ethics, make acts right. (Keller, 2007)

What makes an act, X, right according to virtue ethics? The answer is, of course, that X is right because it is what the virtuous person would do in the circumstances. So, A says that it is always acceptable to be motivated by the consideration that X is what the virtuous person would characteristically do. In addition, Keller argues that A implies the following,

B: It is never undesirable for an agent to be moved to action by the thought that her act is in accordance with the virtues, or by the thought that she is acting as the fully virtuous person would act. (Keller, 2007)
We will come back to the question of whether avoiding self-effacement really requires meeting these two conditions. For now, let’s move to the question of whether virtue ethics is, in fact, self-effacing. Keller argues it is by using the following example.

Arthur, Benjamin and Christine, are spending a night together in the woods, sharing the only hut in an isolated camping ground. It is cold, wet and windy. Peering out through the hut’s window, they notice a family of hikers. The parents are struggling to put up a tent in the wind, and the teary children are eating dry biscuits...Arthur, Benjamin and Christine each decide to invite the family into the hut, to share the food…and to have somewhere warm and dry to sleep…Arthur, Benjamin and Christine have different motivations. Arthur’s primary motivation is to help out the hikers and relieve their misery. “They are so cold and tired and hungry”, he thinks, “and they’ll be much happier in here.” Benjamin’s primary motive is to act generously. If he gives an honest answer to the question of why he wants to ask the family in, he will say, “I want to act generously, and its clear in this case that if I am to be generous then I will invite the hikers to share our hut”...Christine has yet another motivational thought. She wants to do what the fully virtuous person would do. She has a reasonably well-formed notion of the fully virtuous person in mind, and...she wants to emulate that person. She will say, if she is honest, “It seems pretty clear to me that a fully virtuous person would, in this sort of situation, invite the family into the hut. And I want to do what the virtuous person would do.” (Keller, 2007)

Keller takes this example to show clearly that, of the three, Arthur is the most generous and the one who most resembles the fully virtuous person. Why? Because the truly generous motive in the situation is the motive of helping others, even where doing so requires sacrifices. Acting from the thought ‘This would be virtuous’ need not be seriously problematic -- the motive can be to be genuinely virtuous. The problem, according to Keller, is that Benjamin and Christine are lacking in full virtue, and their misdirected thoughts reveal this.

How does this show that virtue ethics is self-effacing? According to Keller, the reason why Arthur more fully approaches virtue is that Arthur acts as the virtuous person would in a robust sense. All three act as the v-person would in a reduced sense -- they all invite the family to share the hut. However, virtue ethics also takes into consideration the internal aspects of action: one’s motives and emotional responses. Aristotle, for example, is clear that virtue or excellence requires that we not only do the right act in a reduced sense (invite the family in) but that we act in the right way. Only Arthur seems to understand this aspect of generosity. Arthur does it from the proper motive, and we might add with the right emotional responses, etc.

Furthermore, this is not a fluke. Generosity is not a special virtue in this respect. For example, modest people do not act from the thought ‘This would

---

1For Aristotle’s claim that focuses particularly on emotion see EN 1106b18-23
Should we believe our Ethical Theories?
Virtue Ethics and Self-effacement

be the modest thing to do,’ only the falsely modest do that. Keller argues that this holds for just about every virtue, with the exception of Justice.

Since the virtue ethics account of right action claims that what makes the act right is that it is what the virtuous person would characteristically do in the circumstances, and this is a thought we cannot, in the vast majority of cases, have in mind when we act virtuously, Keller concludes that virtue ethics is self-effacing.

Rosalind Hursthouse (1999) has helpfully pointed out that when we offer the general schema for a type of moral theory, answering detailed questions about how it guides action or what it says about motivation must await the further spelling out of a particular version of the theory. So, we must hold off on certain questions until someone has filled in the schema with a live theory. Virtue ethics, understood as a classification of a type of theory, does not have a non-schematic account of right action. This is a recognizable way to think about how we develop our moral theories.

For example, consequentialist theories fall under the classification ‘consequentialist’ because their accounts of right action are instances of the following schema,

\[ E: \text{An action is right iff it promotes the best consequences.} \]

Of course, we don’t have a live theory yet. In order to have a live theory, we need to fill out this schema with some understanding of what counts as the best consequences. So, a utilitarian might say,

\[ U: \text{The best consequences are those in which happiness is maximized.} \]

At this point, we are well on our way towards utilitarianism. The same process is required for deontological theories and virtue ethical theories.

If we forget this aspect of theory construction we risk making superficial objections to schemas and fail to make deeper contact with the theory. A more developed account of the virtuous person is needed. So, what does that more developed account look like?

Following Aristotle’s lead, virtuous people do certain sorts of acts. They tell the truth, they help other people, they face danger, they refrain from over-consumption, etc. In addition, they do these things knowingly. Also a virtuous person is not doing what she’s doing simply from impulse. This condition needs to be expanded to include the important idea that she is doing it for the correct reason, not any old reason. Finally, the virtuous person feels the right way when she acts. This is to say that she does not encounter internal resistance when acting for the right reason (e.g. in the form of a strong desire to lie or strong aversion to telling the truth). Going beyond the schema in this way allows us to develop a response to the self-effacingness objection.

Consider the case of Arthur. Of the three agents in the story, Arthur seems to exhibit full virtue more than the others, because Arthur acts for the right reason. The virtuous reason for helping the family is something like ‘They are cold and tired and hungry and would be much better off if they were in here with us.’ Keller takes the theory to be self-effacing because in this situation, and in almost all other situations, the virtuous agent will not have the thought,
at the moment of action, ‘This is the way the virtuous person would act.’ However, no virtue ethicist claims that one should have this thought in mind at the moment of action.

Consider Aristotle’s formulation of what sort of thought the virtuous person has in mind when acting. In EN II.4, Aristotle, focusing on the condition that the virtuous person acts for a reason, articulates what that reason is. The virtuous person acts \textit{proaireoumenos kai proaireoumenos di’ auto} – ‘choosing and choosing for its own sake.’ What is meant by ‘for its own sake’ here?

Bernard Williams (1993) and Hursthouse (1999) have both argued that ‘for its own sake’ should be understood to say that the virtuous person chooses virtuous actions for at least one of a certain range of reasons, where that range is understood by reference to the virtue in question.

So, for example, we can say that courageous people do courageous things ‘for their own sake’ in the sense that courageous people act for reasons like ‘I could probably save him if I climbed up there’ ‘If someone doesn’t go now, they’ll be stuck’ ‘We should speak for those who can’t speak for themselves.’

This yields the following, more complete virtue ethical account of right action, an action is right if it is what the V-person would characteristically do in the circumstances -- and that means doing characteristic kinds of acts knowingly, with the appropriate feeling and, importantly, doing it for one of a range of reasons, where that range is understood by reference to the virtue in question.

This formulation brings to light the fact that the virtue ethics account of right action Keller uses to get his argument going is only a schematized version. But, it suggests more than that. It points to the fact that the account of right action stands for aspects of individual right actions that are the right making features of actions. For example, it was done knowingly and for this reason (selected from amongst the appropriate range of reasons). Call this the \textit{de re} reading of the account of right action.

Once we recognize, as many philosophers have, that an account of right action can be read \textit{de re}, and not \textit{de dicto}, we can see why Keller’s objection misfires. Look back at the camping example. In some sense, though we cannot be sure, Arthur acts virtuously. Of course, we need to know a lot more about the case to make this judgment.

Does Arthur experience internal resistance to helping the family, does Arthur know what he is doing, and, most importantly, is Arthur acting in character? Assuming he is, then he is acting as the fully virtuous person would act, even though he does not have the virtue ethical account of right action in mind at the moment of action.

One might object. Doesn’t this concede Keller’s point? The answer is ‘No’ and should be clear if we consider the difference between a theory’s being self-effacing (either modestly or immodestly) and a theory guiding indirectly. It has long been known that a theory can advocate a two-level view of deliberation, where action is guided by indirection (recall Bob, our music loving tennis player).

It has also long been known that advocating such a view of moral deliberation is problematic when the two levels conflict. Neither Stocker nor those who have developed his arguments ever claimed that the self-effacing
objection arose from the mere fact of indirection. Recall that the guidance Bob’s theory offers is not problematic on the grounds that it is self-effacing. If his theory is problematic, its likely due to the fact that its a crude version of hedonism. But the mere fact of indirection is not, in itself, objectionable.

Now consider how a virtue ethicist would account for our intuitions about Arthur. Arthur can be doing the virtuous thing and not have in mind the virtue ethical account of right action because the virtue ethical account of right action is something that often plays a role at a level of deliberation different from the level that is in play when in the grip of a pressing, real life moral situation. It will only follow that the theory is self-effacing if there is some conflict between the reasons/motivation the theory tells me to adopt and the account of right action. So, is there such a conflict?

A possible test to check this is to ask whether Arthur can hold the following thoughts in his head without conflict, ‘I’m helping the family because they are cold and hungry and will be much better off in here with us.’ ‘Helping the family is the generous (or compassionate or virtuous) thing to do.’

To see that these two thoughts are not in tension, consider a fourth camper, Earl, who is in the hut with Bob and the rest. Earl has been hiking and four-wheeling all day. He longs to eat that warm dinner by the fire. Now this family is outside getting in the way of Earl’s good time. Earl asks Arthur ‘Why are you helping them?’ Arthur responds, as we would expect, ‘Because they are cold and hungry, etc.’ Now imagine that Earl replies ‘So what? Why should I care about their suffering?’ It is perfectly consistent for Arthur to respond ‘It would be unkind of us to sit here, while they are suffering outside.’ Thoughts of the virtues are perfectly consistent with acting from typically virtuous motivations, even if they should not be the immediate motive for virtuous action. Because he mistakenly characterized the virtue ethical, two-level account of moral deliberation as ‘self-effacing’ Keller concluded that virtue ethics suffers from the same problem that Stocker claimed utilitarianism and versions of kantianism do. But this is unsupported. Keller needs a different argument to establish that there really is a worry about self-effacement here.

One could press this issue. Keller can respond by saying that I have pulled a fast one here. Assuming Arthur to be fully virtuous, I have not shown that his reason for action does not conflict with his motive. Surely, we are not justified in taking the action to be right because it is generous or what the virtuous person would do.

This is where things get tricky. Certainly, the action is right because it helps the family, but that’s not the only aspect of the action. Consider Earl again. Imagine Earl says, ‘Alright, I’ll help them come in, but I’m not gonna be nice to them.’ Earl is, in some sense, doing what the virtuous person would do. Of course, in some sense he is not. Earl-style helping is not fully virtuous.

In any case, it should be clear that this objection is somewhat off point. This worry raises the question about the truth or the justification for holding the virtue ethical account of right action. Keller’s objection is that if we assume the virtue ethical account of right action, we get the odd result that the theory is self-effacing. But we can now see that that’s not so.
I have shown how the objection that virtue ethics is self-effacing only works if one ignores the important difference between indirect guidance and self-effacingness. In addition, by reflecting on the two-level theory of deliberation virtue ethics exploits, we see clearly the problems associated with assuming that an account of right action can be used as a guide in particular instances. Of course, more needs to be said concerning how a virtue ethical theory will make a practical difference in one’s life. But, showing that virtue ethics is not self-effacing should remove one hurdle in the way of this further, more positive project in support of a virtue ethical theory.

Bibliography

In this essay I would like to address a substantive challenge to any Virtue Ethics of Care-based conception of the physician-patient relationship. Traditionally the relationship, as well as the normative understanding of the entire medical profession, has been defined in terms of traditional deontic ethical systems. It is a historical fact that since the days of Hippocrates physicians have viewed their behavior toward patients and peers as primarily oath or duty directed. This fact becomes even more apparent when one considers the differences between the physician-patient relationship (and most traditional conceptions of professional relationships) and the relationships one has with family and friends, which are paradigm examples of caring relationships. As a result, attempts to offer a normative account of the physician-patient relationship from a purely aretaic perspective have been criticized for being inadequate, at some point such views must defer to principles, duties, obligations and oaths in order to generate a functional account.

I intend to demonstrate that a stand-alone aretaic account of the physician-patient relationship can be articulated and that such an account possesses all of the advantages claimed to be operative in traditional formulations. Generating a purely aretaic formulation is no easy task since the traditional conceptions of the relationship have always been presented in deontic language, such as the duty of beneficence, the obligation to respect patient autonomy and the oath of confidentiality. In order to be consistent, the Virtue Ethics of Care must explain these apparent deontic features in ethical caring terms and construe the relationship in a completely different fashion; one that resembles more clear-cut caring relationships. I will begin by quickly highlighting the main features of the Virtue Ethics of Care. I will then proceed to describe an aretaic formulation of the physician-patient relationship that will accommodate all of
the desired, and seemingly deontic, features including respect for patient autonomy. This initial description is purely aretaic, making no reference to oaths, duties, principles or obligations. However I will next discuss how such terms can be incorporated into my aretaic account without leading to the charge of inadequacy. I will conclude by addressing a practical criticism charging that any aretaic formulation of the relationship would be incompatible with current medical practice.

The Virtue Ethics of Care

In recent years, a new understanding of the Ethics of Care as a form of Virtue Ethics has emerged (Slote, 2001, 2007, Noddings, 1984, Manning, 1992). So understood, the Virtue Ethics of Care differs from more traditional deontological and teleological views in two substantial ways. First, the focus of Virtue Ethics lies heavily upon the moral agent; the inner traits, dispositions, and motives that qualify her as being virtuous. In contrast, most traditional theories focus primarily on ethical principles, obligations, and/or consequences. The second major difference is the moral language used by Virtue Ethicists when making a moral judgment. Traditional theories tend to judge cases using deontic terms such as “permissible” or “obligatory.” In contrast, virtue ethical views speak of actions in aretaic terms, such as “virtuous” or “admirable.”

To highlight the main features of the Virtue Ethics of Care I will begin by presenting the following moral standard: An action, or person, is praiseworthy or admirable if it, or he, displays a motive of caring or concern shaped by sympathetic understanding, sensitivity to context, and balanced care. Sympathetic understanding occurs when the moral agent makes a genuine effort to come to understand the perspective of the recipient/s of his caring efforts. This is best achieved through open and honest two-way communication. The moral agent also creates and fosters a desire to serve the expressed and genuine needs of the cared-for. This desire to serve or promote the welfare of the cared-for is also shaped by the caring agent’s sensitivity to context. Here, the moral agent formulates plans of action on behalf of the cared-for by prudently considering contextual details surrounding the particular situation and the particular persons involved. One set of details to be considered is the relationship(s) the agent has with the persons involved. This includes the rather general relationship one has with fellow human beings (perhaps even all living creatures) as well as specific professional or role relationships. It also includes an awareness of one’s own talents, abilities and limitations. Finally, the admirable caring agent is influenced by balanced care. Generally speaking, the moral agent must balance her efforts and resources among three general classes of people. She displays intimate concern for those near and dear, humanitarian concern for distant others, and special concern for herself. The distribution of efforts and resources need not be equal. However, in specific contexts balanced care requires that no group of persons, or individual bearing on a particular context may be utterly ignored in favor of some other group of persons or individual.
The standard described above generates a continuum of moral judgments of acts and agents from the Virtue Ethics of Care perspective. At one extreme are acts that are most admirable or praiseworthy, shifting over to acts or agents that are less admirable (perhaps lacking sensitivity to context), to acts that are neither admirable nor deplorable, and finally to acts that are so antithetic to ethical caring (motivated by malice, for example) that they are utterly deplorable. Armed with this conception of ethical caring I will now proceed to describe a new normative account of the physician-patient relationship.

The Areataic Formulation

A purely aretaic normative account of the physician-patient relationship must assert that physicians need not be governed by oaths, duties and/or obligations at all. In such a case the relationship will be driven by ethical caring with substantial aid from a clear understanding of one’s capabilities and the role one assumes when becoming a physician. Awareness of her abilities (and limitations) will help to ensure that a physician will not harm and will do her best to promote the welfare of a patient. Any caring agent, health care professional or otherwise, will act in such a way that will bring all of her talents, skills, and abilities to bear on the situation. To hold back one’s talents and resources for reasons such as selfishness, malice, or lack of interest, displays an unacceptable lack of concern for the cared-for. For example, an ER physician who does a half-hearted or overly hasty assessment and treatment of a walk-in patient (perhaps to get to the golf greens faster) is unethical not for breaking an oath or ignoring a duty, but for failing to bring all of his personal medical resources and expertise to bear in caring for the patient. This argument is in line with the common sense belief that the person who can swim is more to blame, morally, for failing to rescue a drowning child than one who cannot swim. It is the skills, training, and experience of the agent in conjunction with the context that determines the degree of moral culpability. The swimmer is more capable, in virtue of his skill, of saving the drowning child than the non-swimmer. By analogy, a trained health care professional has a greater responsibility toward the sick than the rest of us.

While my discussion of capabilities could supply a care-based justification for traditionally recognized duties of beneficence and nonmaleficence, I still need to offer alternatives to other seemingly deontic elements that have traditionally shaped the physician-patient relationship. I am referring to duties concerning respect for autonomy, truthfulness, confidentiality, and justice.

Traditionally, the concept of autonomy has been presented in deontic terms. Some, like Thomas Hill, have construed autonomy as a right possessed by all persons and to be honored by all persons (Hill, 1987). Others, like Beauchamp and Childress, have safeguarded autonomy by incorporating it into a principle to guide physician conduct (Beauchamp and Childress, 1994, Ch. 3). Neither of these conceptions will be adequate in a purely aretaic formulation of the physician-patient relationship. In order to be compatible with a Virtue Ethics of Care, autonomy must accommodate connectedness, be sensitive to context, and be presented in non-deontic terms.
These objectives can be best achieved by understanding the notion of autonomy in two important ways. First autonomy must be understood to be a vital component of human welfare that all individuals, including physicians, must be concerned about. Second, autonomy must be understood as a central component of caring relationships. Under most conditions a caring relationship involves two or more agents who communicate with each other in open two-way dialogue. The one-caring, wishing to promote the welfare of the cared-for, attempts to come to understand, via sympathetic understanding, the fears, wants, desires and plans of the cared-for so that he may determine appropriate action. Under optimal conditions, this is best achieved when the cared-for is a fully autonomous agent; she freely exchanges information with the one-caring.

Under this formulation, autonomy becomes something not merely to be respected or honored but rather an element of human welfare and relationships to be nurtured, restored, and empowered. In the context of medicine, a doctor’s concern for patient autonomy can manifest itself in a number of ways. First the doctor can contribute greatly to a patient’s deliberative process by openly and honestly communicating relevant information. Second, in order to show concern for this part of the patient’s welfare, a doctor must also be sensitive to the particular context surrounding the patient at the particular time of care. This can be achieved through sympathetic understanding. For example, in order to best communicate sophisticated medical information clearly and efficiently a physician might first engage in some kind of dialogue with the patient to determine his level of understanding. This hopefully will help the physician to convey the information at a level that will be grasped by the patient. Finally, in cases where a patient’s deliberative abilities are questionable, constrained by physical or external influences, or outright out of commission (due to unconsciousness, for example), the doctor should act to restore those abilities as quickly and as best as possible.

With regard to apparent deontic notions of justice, truthfulness and confidentiality, I believe that appeal to conventional roles will be helpful. While it is certainly possible to develop an elaborate, role-based morality here, it will be sufficient in this context to confine the discussion to a basic conception of roles. In addition, it is important to note that I will be discussing roles in a way that will be similar to how I discuss duties, oaths and obligations in the next section. Roles must be understood as instrumental aids to ethical caring only, not as having independent moral status, and not as being prime constituents of the relationship.

To begin, I think that at a common sense level most roles are defined by the function or purpose they serve in a culture. This certainly seems to be the case with parenting. The role of parent is generally defined as involving the provision, support, and promotion of the personal development and general welfare of a child. The role of a physician can be defined in a similar fashion, by identifying what Pelligrino and Thomasma call the ends of medicine. They offer the following description which is adequate for purposes of my discussion:

---

Briefly, the ends of medicine are ultimately the restoration and improvement of health and, more proximately, to heal, that is to cure illness and disease or, when this is not possible, to care for and help the patient to live with residual pain, discomfort, or disability (Pelligrino and Thomasma, 1993, pp. 52-53).

They also note that “health” should be given a broad interpretation that goes beyond mere physical conditions to include psychological health as well. This should not suggest that physicians should be concerned about mental health to the degree that psychologists are. As mentioned earlier, they must remain constantly aware of their abilities and limitations. It does mean that they should be concerned with more than mere treatment of physical ailments. I have already noted that the genuine need to make autonomous decisions must be given consideration by the physician. This will often involve not only efforts to restore a patient’s physical capabilities but also, through communication, the restoration of self-confidence, and the provision of relevant information in order to make informed and competent decisions.

The broad construal of the ends of medicine could arguably pave the way to associating truthfulness and the maintaining of confidentiality with the role of physician. There is no question these days that psychological health and physical health are intimately related to each other. If truthfulness and confidentiality are important to one’s psychological health and that, in turn, bears on one’s physical health, it does not seem a stretch to include such elements into the role of a physician. I admit that this is a questionable empirical claim and will only offer some brief support here.

First, honesty among physicians is necessary in order to enable and empower patients to make informed decisions. If autonomy is to be respected, it seems that a physician must also be truthful so that autonomous decisions can be made. The relevance of respect for confidentiality in considering the psychological health of most patients can by justified by noting the special circumstances surrounding the physician-patient relationship. The kind of information a patient shares with a health care professional is frequently of a deeply intimate and private nature. For many patients, peace of mind is secured when they are confident that such information will remain private. I am not claiming that truthfulness and confidentiality are sacrosanct, just that they are psychological needs that are important enough to warrant their inclusion in the role of physician.

I still have not discussed how the notion of justice is associated with the role of physician. To do this one needs to look at a specific element common to a number of different roles. Again, at a common sense level, one thing that many different roles—such as that of parent, priest, doctor, lawyer, and judge—have in common is a fiduciary aspect. Though the details of the fiduciary element vary from role to role, I think there is a common assumption present in all. That assumption is that whatever welfare the fiduciary is protecting or promoting, be it physical health, spiritual health, legal expertise, etc., she will do it fairly, i.e., be directed by a sense of justice. Take the example of parents with more than one child in their care. In promoting the
development and welfare of their children, with rewards and, at times, punishment for example, it is assumed that parent’s efforts will be distributed fairly among all the children. I believe this conventional assumption carries over to other fiduciary roles, including that of physicians. In delivering health care, physicians should not arbitrarily discriminate.

Here it will be helpful to say a few things about the creation of professional roles. In order to be in line with the Virtue Ethics of Care it must be asserted that the creation and development of professional roles, while largely conventional, must be explained in terms of ethical caring. I mentioned earlier that a role is commonly defined by the function or purpose it serves in a culture. In addition, I would suggest that the creation of conventional professional roles is motivated by concern for the recipients of the profession’s services, members of the community, that they feel safe and secure about seeking professional services. This is particularly the case with professions whose services have great social worth, as is the case with the health care profession, or those with greater social/political power, such as law enforcement.

Rather than being driven by oaths, duties, and obligations, a physician should enter into a relationship with a patient shaped by ethical caring which is in turn constrained by the role specific elements of the medical profession. He accepts such a role because it has been established by the profession itself and has been acknowledged by the community. To ignore or reject the conventional understanding of the role, or to be selective as to what elements of the role one will be faithful to, is to display a lack of concern toward others. It shows a lack of caring toward the recipients of the physician’s services since they enter the relationship assuming that the role has been fully embraced. A caring physician must be concerned about the public’s views, perceptions, and assumptions regarding the profession. Being true to the role is one way to do this.¹

Now it might be argued that the assumption of a role, particularly a professional one, carries with it some kind of implicit promise or oath. If that is true, then I am right back where I started. But I do not believe it is necessary to understand roles in this way. It is arguable that the deliberate assumption of roles such as parent, family member, friend, or even mentor does not carry implicit promises with them. If this is true it does not seem implausible to think of professional roles in a similar spirit. By re-conceiving the driving force of the physician-patient relationship in terms of ethical caring guided by roles I hope to have held true to one of the prevailing themes of care-based ethics, specifically the emphasis on relationships and connectedness between persons.

An Aretaic Account of Oaths, Duties, and Obligations

While it can be argued that the aretaic account provided above can stand alone as a normative account of the physician-patient relationship, it is prudent

¹It could also be argued that failing to fully assume a professional role displays a lack of concern toward the profession itself since it implies that one does not care about the conventions established by the profession.
to discuss how the Virtue Ethics of Care can accommodate the traditional deontic features of the relationship. This would present the advantage of keeping those features in place while operating from within the Virtue Ethics of Care framework. However, it is vital to note at the onset that such features must be understood as instruments of ethical caring and not as primary constituents of the relationship.

Here there is help from both Rita Manning and Michael Slote. In Speaking From the Heart, Manning offers an account of ethical caring that accommodates the use of rules and rights (Manning, 1992, p. 74). She states that if they arise from genuine ethical concern for persons, rights and rules can be used to establish a minimum level of acceptable behavior and are particularly useful in guiding actions toward strangers or caring for several persons at (roughly) the same time.

This line of thinking could arguably be extended to include the establishment of oaths, duties, and obligations for physicians. They are put in place out of genuine concern for patients and serve as a measure (establishing the minimum) of caring behavior. So, for example, no matter what they do in caring for a patient, health care professionals must at least be committed to promoting health, to not causing harm, to being just, and to maintaining confidentiality. Again, such oaths and acting in accordance with them are understood to be aids to caring. Their presence and influence should in no way diminish a health care professional’s genuine concern, shaped by engrossment, sensitivity to context, and balancing, for the welfare of their patients.

In his most recent work, Slote offers his own caring-based account of apparently deontological behavior, like oath or promise keeping, which could also be appealed to for an understanding the physician-patient relationship (Slote, 2007, 45-53). Put crudely, Slote argues that the moral significance of promise keeping can be established by reference to empathic concern for the welfare of others. Looking to the work of T. M. Scanlon, Slote notes that the wrongfulness of promise breaking lies in the disappointment of induced expectations created when the promise is made (T. M. Scanlon, 1998, Ch. 7). He breaks from Scanlon in accounting for that wrongness by appealing to empathic caring rather than deontic terms, obligations, duties, and constraints.

Slote asks us to consider two moral agents. The first agent makes a promise to another person only to break it at a later time. This imposes a certain degree of harm on the promisee. The second agent is a bystander who could aid the promisee by fulfilling what was promised by the first agent, but ultimately decides against such action. The morally significant difference between the two agents lies in the degree of causal connection to the resulting harm each has. Obviously, the first agent plays a much more substantial causal role in creating the harm than the bystander because he induced an expectation through his promise and then disappointed it through his lack of appropriate action. Slote observes that from an empathic standpoint, we would choose not to establish that kind of causal connection whenever possible. In other words, a caring agent will be wary of making promises she cannot keep and will certainly not make ones she has no intention of keeping due to the potential (or guaranteed) harm it will create by disappointing the expectations of the promise (Slote, 2007, 46-47).
This conception carries over to the medical context nicely. A health care professional’s formal oath certainly generates a set of expectations among patients (i.e., that they will not be harmed, that information will be kept confidential, etc.). To take such an oath with no intention of honoring it, or even with some degree of uncertainty or lack of conviction, would clearly display a lack of concern for the welfare of patients who will fall under the oath taker’s care.

Slote’s account of promises lends additional support to my earlier discussion of the physician’s assumption of a professional role. Much like promises, the role/s one assumes can induce expectations among those with whom one enters a relationship. The specific kinds of expectations induced are determined by the role/s. For example, a patient develops certain expectations (to be aided, to be treated fairly, to have personal information be kept confidential, etc.) once they visit a doctor because of the role he has assumed. Therefore, a doctor, or any professional, parent, mentor, or even friend, must take the role/s they assume very seriously. To fail to do so would display a lack of ethical concern for those whose expectations may be disappointed.

Even if the aretaic construal of the physician-patient relationship is possible, the realities of modern medical practice make it prudent to say something about the traditional physician or health care professional who insists upon conducting himself by deference to only duties and principles. Obviously, from the Virtue Ethics of Care perspective, this is not ethically optimal behavior. As Slote observed, a person who first defers to duties instead of displaying immediate and direct concern for the welfare of others takes a step away from the connectedness that is basic to ethical caring (Slote, 2001, p. 49). Still, a person motivated by conscientiousness could still warrant some praise compared to one who is motivated purely by malice and/or self-interest. It is important to note here that knowing the specific duties or principles one is deferring to will be critical in assessing praise since there are many duties that will display no concern at all for the welfare of those involved. Slote gives an example of a conscientious Nazi concentration camp guard who executes prisoners out of a sense of duty to the Reich (Ibid., 51-53). So, if we are going to praise conscientious physicians on occasion, it is likely that they will be acting in accordance with principles such as those offered by Beauchamp and Childress. Again, this is not optimal and a doctor’s motives toward the welfare of the cared-for will always be the ultimate basis for praise or blame. It is quite conceivable that a conscientious physician, motivated by the duty of benevolence, will show a lack of concern for a patient, for example, by ignoring the patient’s desire to be involved in the decision-making process.

**Conclusion: Practical Matters**

Even if I have succeeded in developing a purely aretaic account of the physician-patient relationship, critics will likely ask, ‘why bother?’ After all, the traditional construal has been around for some time and it fits nicely with modern medicine. I will devote the remaining paragraphs of this essay to offering a response to this practical problem. To do so I will first suggest that
the Virtue Ethics of Care can meet an as-of-yet unfulfilled desire of patients in today’s health care system. I will finish by pointing to the comments of physicians who have actually advocated ethical caring for patients.

A number of critics of ethical caring in medical practice make reference to the realities of modern medical practice. Robert Veatch referred to it as ‘stranger medicine’ in light of the shift to specialization with its emphasis on technique, skill, and technology (Veatch, 1985). I have argued elsewhere that simply because stranger medicine is a reality does not mean it cannot be ethically challenged or that an ethical theory like the Virtue Ethics of Care must accommodate it or risk rejection (Matthew McCabe, ‘In Defense of Ethically Caring Physicians,’ Work in Progress). Daniel Putman has also responded to Veatch (Putman, 1988). Putman notes that stranger medicine has created an undesirable condition among patients. The shift toward specialization and technical skill among doctors has resulted in a shift away from communication and interpersonal skills. This has left patients with a growing feeling of separation and isolation, compounding their vulnerability in their relations with health care professionals. In another work, The Crisis of Care, E. Dawn Swaby-Ellis (a physician) cites a 1992 American Medical Society survey that reflects patient dissatisfaction with doctors (Phillips and Brenner, 1994, p. 84). The results indicate that less than half the American public believes that ‘doctors usually explain things well to their patients,’ three fourths think that ‘doctors keep patients waiting too long,’ two thirds think that ‘doctors are too interested in making money,’ and six out of ten believe that ‘doctors don’t care about people as much as they used to.’

If patient dissatisfaction with distant, oath-driven physicians is a real problem then I think the aretaic-based Virtue Ethics of Care deserves consideration as an alternative to ‘stranger medicine.’ If patients desire greater connection with their physicians, the ethical responsibilities it attaches to physicians seems uniquely suited to meeting such a need. I am not suggesting a return to the days of ‘old Doc Smith,’ a dear friend of the family. I am simply claiming that even specialists can conduct themselves with ethical concern toward their patients as well as utilize their special talents. Specialization and ethical caring are not, or need not be, mutually exclusive.

One might respond by asserting that such a shift would be bad for the medical profession and ultimately bad for patients. However, there are some indications that members of the medical profession actually want to incorporate ethical caring in today’s health care system. Swaby-Ellis argues, in The Crisis of Care for the caring physician who must balance effectiveness, efficiency, and empathy (Phillips and Brenner, 1994, p. 83-94). In another text, Empathy and the Practice of Medicine, a number of physicians and other health care professionals, led by Howard Spiro, voice their support for a kind of conduct in line with ethical caring. Spiro makes the following observations that seem to come right out of the Virtue Ethics of Care.

Empathy, however, underlies the qualities of the humanistic physician and should frame the skills of all professionals who care for patients (Spiro, et. al., 1993, p. 7).
Physicians are more than conduits of pills and procedures and they should be more than transparent technicians; to quote the old phrase, they need character (Ibid., pp. 11-12).

So, one final response to the critic’s pragmatic concern about ethical caring in today’s medical practice is that it is actually valued and desired by members of the profession.

I hope to have demonstrated that a Virtue Ethics of Care can supply a workable foundation to a normative account of the physician-patient relationship. Contrary to the concerns of critics, it need not eliminate or ignore such crucial elements as autonomy, justice, confidentiality, etc. In addition, it can offer an account of traditional deontic features, such as oaths, duties and obligations, in aretaic terms if necessary. Lastly, I hope to have shown that there is at least some indication that such a formulation has been and would be endorsed by patients and physicians in current medical practice.

References


One of the most challenging types of problem for any ethical theory involves the specification of an individual’s moral obligations in what I will refer to as collective action situations. Global warming, world hunger, and disease control and eradication are examples of this sort of problem. There are six morally salient aspects of these situations. First, something that is very bad for many people either is occurring or will occur unless some action is taken. Second, an individual cannot completely prevent or eliminate the harm from occurring no matter how much he gives or sacrifices, nor can he individually make a significant contribution to eradicating or solving the problem. Third, if everyone who could contribute to preventing the harm from occurring were to do so, it would not occur—collective or group action could solve the problem. Fourth, if this collective action were to occur, the cost for each of the individual agents would be minimal. Fifth, those who are in a position to help alleviate or prevent the harmful situation from occurring are largely distinct from those who will be harmed or whose situation will be significantly worsened—the situation requires altruistic behavior. And sixth, there is a high degree of uncertainty surrounding these situations. It is very difficult to determine the likelihood of how much a person will contribute and how many potential contributors will actually contribute. Exactly how much time, money, or effort it will take to avert the problem is unknown. The number of people who will be harmed and the degree to which they will be harmed can only be estimated. These questions may be tackled empirically and while some answers will have more support and plausibility than others, it is very unlikely that we can ever approach anything like knowledge in our answers to them.
Peter Singer’s well-known article on hunger exemplifies the utilitarian approach to such cases. His conclusion is that individuals are morally obligated to do as much as they can to prevent or alleviate harm, even if this requires a fairly heroic level of sacrifice. Anything less amounts to a failure to meet one’s obligations, so one acts wrongly if one gives anything below this significant level. I will refer to this as the heroic sacrifice principle. Given that act utilitarianism requires such a high level of sacrifice, and given that many people have the intuition that we cannot be obligated to engage in this level of sacrifice, Singer’s argument can be seen as a reductio of act utilitarianism rather than as establishing a duty to do all we can to alleviate hunger. The criticism of the heroic sacrifice principle has its roots in the question of how demanding moral principles can be. As moral requirements become more demanding and restrictive, and especially as they limit or curtail our ability to engage in other innocent or valuable pursuits such as producing art or contributing to the stock of human knowledge or spending pleasurable time with family and friends, their status as genuine moral principles can be questioned. Moral rules which are as restrictive as is the principle of heroic sacrifice might, in the extreme, be said to mandate a kind of enslavement to morality. I take it that this worry, or something like it, is at the heart of the worries expressed by utilitarianism’s critics, from thinkers as diverse as Nietzsche and Bernard Williams.

But then what sort of moral principle does govern these situations of collective action? What sort of obligation do those of us who can help alleviate these problems have? I want to suggest here that the sort of contractualist theory laid out by T.M. Scanlon provides a helpful framework for thinking about collective action cases and for understanding and justifying whatever principles turn out to be applicable to these cases.

Scanlon’s Theory

Scanlon’s thesis is that the correctness or acceptability of a proposed moral principle is to be determined by asking whether or not the principle can be reasonably rejected. Here is one formulation of his view, ‘judgments of right and wrong...are judgments about what would be permitted by principles that could not reasonably be rejected by people who were moved to find principles for the general regulation of behavior that others, similarly motivated, could not reasonably reject. In particular, an act is wrong if and only if any principle that permitted it would be one that could reasonably be rejected by people with the motivation just described (or, equivalently, if and only if it would be disallowed by any principle that such people could not reasonably reject).’

The basic idea is that when people advance competing claims about what sort of treatment they are owed or how they or others are obligated to behave,
Utilitarianism, Obligation, and Significant Sacrifice

we need to look at various rules which might be proposed to settle the issue. A perfect rule which works to everyone’s benefit or which will satisfy everyone and address their concerns while imposing no burdens on anyone is unlikely to be found. But presumably there will be a rule (or rules) which is the best of a number of imperfect solutions. What makes one rule superior to the others is that there are stronger, better reasons for rejecting all the other proposed principles. The idea of reasonable rejectability is the determining factor here. Moral rules are those that cannot be rejected by a reasonable person, even if he knows or suspects that he will be burdened by the rule, or that certain activities which he wants to engage in will be curtailed, since for any other rule which he proposes which will permit him to act the way he desires or to avoid the burdens, there will better reasons for rejecting those than can be advanced for rejecting the actual rule which imposes those costs on him. The rules which are not reasonably rejectable form the basic framework of ethics.

A couple of points concerning this view should be mentioned to avoid any misunderstanding about the sorts of reasons that can be advanced against a principle, including those that might be proposed for collective action situations. First, Scanlon sharply distinguishes between morality in the narrow sense, which asks what we owe one another, and morality in the broader sense, which concerns itself with the question of value and more broadly of what we have reason to do with respect to objects and pursuits which are deemed valuable. One implication of this distinction is that there are potentially two different kinds of reasons for rejecting a principle, personal and impersonal ones. Impersonal reasons are those reasons that appeal to a thing’s or activity’s value for its own sake. Personal reasons on the other hand are reasons that appeal to a person’s well being, or to the effects or outcome of certain kinds of treatment for an individual, or to whether he will be expected to perform certain kinds of actions which may be burdensome. For example, suppose the owner of a work of art wants to destroy it (perhaps because he has come to believe it is immoral or decadent); but others are horrified by his intention and think he should not be permitted to do that. The owner proposes that we adopt a rule which permits the owner of something to destroy whatever he owns. Now one obvious reason to reject this rule is that if the object in question is a genuine work of art, then the principle will enable things that have significant value to be destroyed at the whim of their owner. That counts as an impersonal reason for rejecting the rule. A personal reason for rejecting the rule might be that if the artwork is destroyed, it will make it impossible for people to view the artwork and gain whatever inspiration or other kind of aesthetic experience is typically aroused by such an object. That is a personal reason.

Scanlon’s claim, which cannot be defended here, is that only personal reasons count when it comes to determining which rules to reject. This is

---

1See Scanlon (2000), pp. 171-177 for a more detailed account of this distinction and some of the issues that it raises.

2Scanlon himself talks about impersonal values (Scanlon (2000), pp. 218-223), but implicit is the distinction between these two kinds of reasons. For a complete account of this distinction and the crucial role it plays in Scanlon’s theory see Kumar (2003). Michael Ridge makes many of the same points when he draws the distinction between agent-relative and agent-neutral reasons; see Ridge (2001).
important, for any rule which tells us how to respond to global warming, for example, cannot be rejected because it will lead to the destruction of supposedly valuable things such as species and ecosystems. The reasons must be reasons from the standpoint of a person, and they must somehow be relevant to persons.

Now having said that, it is important to keep in mind that not all personal reasons are reasons of well-being, at least if well-being is construed narrowly in terms of how one is faring or getting on. For example, that a rule treats someone unfairly is a reason to reject the rule. But being treated unfairly need not affect a person’s well-being. If he is used to being treated that way, he will not find it objectionable or psychologically troubling. Furthermore, unfairness can occur even though everyone’s well-being is increased. Suppose some good which no one has a special claim to is to be divided. All other things being equal, if the division is unequal, then some are being treated unfairly. But no one’s well-being has been diminished, and in fact everyone’s well-being has increased, since everyone now has at least somewhat more of the good than what they had earlier. Thus the objection to the principle which permits an unequal distribution cannot be in terms of the fact that it diminishes one’s well-being or affects it negatively. So the objection must be just that it treats some people unfairly.

**Four Competing Principles**

In cases of collective action, we need to consider reasons relating to those who will be harmed if no action is taken as well as reasons relating to those who will be inconvenienced by having to offer some form of aid. Four alternative principles might be thought to have some initial plausibility and so which might be proposed to govern these cases. In order of placing increasing demands on agents, they are:

1) The principle of non-aid- there is no obligation to help or provide aid in collective action cases (absent special obligations such as having promised to help). A person is always permitted to do nothing and not to provide aid.

2) The principle of proportional sacrifice- individuals are obligated to do or give an amount such that if all relevantly similar individuals were to contribute their proportional share, whether through their actions or otherwise, such as through financial contributions, the harm that would otherwise occur will not occur.

3) The principle of easy rescue- individuals are obligated to do or give an amount such that the cost or burden to the agent is minimal or insubstantial in terms of time, money, or effort.

4) The principle of heroic sacrifice- individuals are obligated to give as much as they can up to the point where they would be made worse off than those to whom the aid is being given.
**Comments on the Various Principles**

Scanlon discusses the easy rescue principle in the context of rescuing individuals, but it can be applied here as well. The idea is that a person must provide a level of aid which, although it will not have a significant impact on his well-being or ability to pursue those activities which he takes to be valuable, will nonetheless require some donation of time, money, or effort. It is difficult to specify exactly how much, but there are all sorts of intuitively obvious applications. Suppose I have just bought a new pair of dress shoes. I am wearing them when I see a toddler drowning in a large wading pond. I have no time to take them off. Obviously the rescue principle obligates me to save the toddler, even though it means ruining my new shoes. On the other hand, donating a kidney to save a stranger might be required by the heroic sacrifice principle but is not required by this one, since donating a kidney is a significant sacrifice. Perhaps what this principle requires is along the same idea as the religious notion of a tithe.

The principle of proportionality generally will require less than the principle of easy rescue, at least in cases of collective action, because typically in these cases there are a number of people who are in a position to help, and thus the burden can be spread over more people, thereby lessening what is required of each individual. It should be noted that unlike the heroic rescue and easy rescue principles, the amount required by this principle is partly a function of how many people are in a position to help— the more people who are able to help, the less each person is required to contribute. In the case of the drowning toddler, for example, if there are ten people who are all in a position to prevent the drowning, one of us will actually have to go in and ruin his shoes, but the other nine would be obligated to chip in to pay for 90% of the ruined shoes, so that the sacrifice of the person who actually saves the child is less than what he would be called on to sacrifice by the principle of easy rescue. According to the principle of easy rescue I am obligated to make the sacrifice regardless of what others do, and others have no obligation to try to make up to me any part of what I sacrificed or donated, even though what I gave up was itself not a substantial sacrifice.

It is also worth emphasizing that the principle of proportionality does not require that everyone contribute exactly the same amount. Rather, it leaves open the question of what determines the level of a person’s contribution, and I suspect that there is no precise answer to this question, since the level of one’s contribution will be determined by various factors. In some cases the major determinant may be ability to contribute— the more of whatever is needed that a person has, the more that is expected of her. In other cases questions of one’s contribution to the problem may be relevant, whether that contribution was intentional or inadvertent, and so forth.

---

1Scanlon discusses both easy rescue cases and cases requiring a more significant sacrifice on pp. 224-229 of Scanlon (2000).
A Thought Experiment

Suppose that a significant harm that will affect a large group of people is about to occur. There are a number of buttons and if these are pressed and held long enough, the harm will be averted. No one is sure how many minutes of button pushing needs to be logged before the harm is averted, but the best guess is that about 12 hours (or 720 minutes) of button pushing will be enough to avert the harm. Conveniently, there are 720 people who are in a position to press and hold one of the buttons. It is impossible to completely co-ordinate the effort, so that no one can be sure whether or for how long everyone else will press a button, although some rough and incomplete data is available. It is impossible for any one person to press any more than one button at a time. There are no transaction costs- it takes neither time nor money to get to a button- they are ubiquitous. All other conditions are equal- no one person is in a better position than another to push a button or has more free time, none of the possible times for pushing buttons are better for any reason than any of the other times, and so forth.

If we apply the principles here, what sort of specific obligation would a person have to push a button, thereby helping to avert the harm? The proportionality principle is the most straightforward. Since there are 720 people in a position to help, and since 720 minutes of button pushing is required, each person is obligated to push the button for one minute.

The other two principles cannot be applied quite as neatly, but we can perhaps provide a rough estimate, assuming normal background conditions for a normal person going through a normal day. The easy rescue principle requires us to perform easy rescues, to not engage in any great sacrifice but rather to contribute something that will help to avert the serious harm while not significantly affecting our own well-being or the quality of our lives. The question then is, how much time in an average day could a normal person give up to spend pushing buttons without sacrificing anything of significant importance or value, including time spent relaxing, which enables a person to accomplish whatever important tasks they want to do during the remainder of their day. Suppose the average person whiles away 20 minutes surfing the net or doing other sorts of things that really contribute nothing significant to their day or to their valuable goals and pursuits or quality of life. Then the easy rescue principle would require that everyone contribute 20 minutes.

The heroic rescue principle requires everyone to make a significant sacrifice. For example, a person might quit his full time job and take a part time one. Perhaps he could spend less time cooking dinner for himself and his family and his family could eat sandwiches every evening. He might cut down on the amount of time he spends reading. The sorts of things the person could achieve would be seriously curtailed, but he certainly could live like that. He would then be able to devote the rest of his time to button pushing. Suppose the average person could contribute 6 hours to button pushing. Then that is what would be required by the heroic rescue principle.

Given these conditions, which principle cannot be reasonably rejected? Clearly the principle of non-aid is highly problematic. Those who are in need of help and who could be helped through a minimal effort on the part of each
individual in a position to help have a very good reason to reject that principle, because if no action is taken they will end up suffering some great harm. So there is a good reason to reject the principle of non-aid. Of course those who will be given the liberty not to assist those in need have no reason to reject the rule and some minimal reason to opt for it, namely they will be left at liberty to do whatever they want to do with their time and money. However, if a rule can be imposed which has minimal costs and which can resolve a serious problem, then those who would suffer have a better reason to reject the rule which requires nothing of others.¹

Now all three of the other principles, if acted on, will prevent the harm from occurring, so none of them can be rejected by those who will be harmed if help is not forthcoming. That is, if everyone acts on their duty and follows any one of the three principles, sufficient aid will be given to avert the harm. Of these three, the principle of proportionality requires the least from donors in terms of aid. Thus it appears that this one is cannot reasonably be rejected by anyone- it requires the least of potential donors while providing sufficient aid to avert the harm that will occur, which means it cannot be rejected by those who are threatened by the harm. Of course if everyone acts on either the principle of easy rescue or the principle of heroic sacrifice, there will be a surplus of aid- perhaps there could be an aid rebate, or perhaps surplus aid gets used up and is essentially lost. But those who are threatened by harm cannot reject the principle of proportionality just because it fails to provide them with this surplus, because the surplus is not needed to avert the harm.

A Problem for the Principle of Proportionality

Suppose however that persons in need of aid offer the following reason for rejecting the principle of proportionality. That principle, although it requires little from each individual, will only be successful at staving off the harm that will occur if there is practically universal compliance. But such a high degree of compliance is unlikely, knowing what we know about people and the likelihood of their providing aid. Thus a rule which requires the bare minimum of aid can be rejected because the total amount of required aid is unlikely to be forthcoming. The rule that is most likely to produce the level of aid required is the one that makes the most stringent demands on donors, and so there is a very good reason to prefer the heroic sacrifice principle to the two other, less demanding principles.

Now this is a serious concern, but there are at least two ways to respond to it. The first is that the claim that a stronger principle is more likely to bring in the required aid than is a weaker principle is not obviously correct. In order to

¹It is worth noting in passing that sometimes the claim is made that the best way to prevent harm from occurring in some cases of collective action is not to introduce outside help and to allow those who are threatened by the harm to work to avert it themselves. I think this sort of claim is almost always empirically false and a way of avoiding responsibility- but if there are any cases where the empirical claim is correct then there will be no good reason to reject the principle of non-aid.
see this we might recur to the thought experiment. Although the three competing principles vary in terms of how much participation is required for them to be successful, they also vary in terms of their demandingness. The proportional principle is the most straightforward - although it requires almost everyone to spend some time pushing buttons, it only requires one minute of each person’s day, an amount of time which is quite insignificant, all other things being equal. Giving up 20 minutes as required by the easy rescue principle is not terribly burdensome - people will have to forego something, but most people can find 20 minutes in their daily schedules when they are not engaged in any kind of significant or valuable activity, although it is more burdensome than one minute. Finally, if people give up 6 hours a day to press buttons, they will be seriously shortchanging themselves and their families in all sorts of ways. So although it is unlikely that everyone will contribute, it is also increasingly unlikely that any given individual will contribute the obligatory amount of aid, as that amount increases. Thus, although it is more likely that two people will contribute as opposed to 720, it is equally unlikely that anyone will give up 6 hours as opposed to just a single minute. My guess is that these odds more or less cancel one another out, and that the problem is equally likely (or unlikely) to be solved regardless of which rule is in place. Additionally, giving more aid than is required would still be considered supererogatory, and some people will give more than the bare minimum that they are obligated to give. If that is correct then there is no good reason to reject any one of the three principles based on its likelihood of success.¹

There is an additional reason for rejecting principles that require people to do more than what would be required of them by the proportionality rule. In collective action cases, whether or not my aid is successful ultimately depends on the actions of others - whether and to what extent they give. This is one of the characteristics of collective action cases - no one individual can prevent the harm from occurring on his own. Given that, no matter how much I give, my giving might ultimately be unsuccessful and a waste of effort. Now any rule can be reasonably rejected that requires a wasted effort. But the more demanding the moral rule that we adopt in such cases, the greater the amount of wasted effort or giving. Thus a rule that sets up the possibility of less wasted effort is always to be preferred to the one that sets up the possibility of greater wasted effort. Thus there is a reason to reject any rule more stringent or demanding than the one which requires the least from individuals while at the same time solving the problem, which is what the principle of proportionality does.

### Individuable and Non-individuable Situations

There is however a distinct objection to the proportionality principle which becomes apparent if we return to the thought experiment. The original version

---

¹Why not then just set the level as high as possible? Because setting such a high standard that admittedly few people could attain would seriously detract from a person’s motivation to be moral, just because morality required such a high standard.
of the thought experiment is intended to model a problem like global warming. A key aspect of that problem is that it non-individuable. The harm that will be caused by global warming cannot be averted for individual persons. Either global warming is rectified by collective action, and the harm occurs to no one, or it fails to be dealt with and the harm occurs to everyone. Thus, unless a given threshold of aid is achieved, no one will be helped.

Problems like hunger and disease control do not work like that- in these cases, the harm is individuable. That is, although an individual cannot solve the problem of world hunger or AIDs, say, through her individual actions, she can improve the situation of at least a few individuals threatened by the harm. Consequently this alters one critical condition of the thought experiment. The buttons still need to be pushed for a total of 12 hours to completely eliminate the problem and prevent the significant harm from occurring. But now, for every couple of minutes that a button is pushed, it is possible to ensure that some unspecified individual will not suffer harm.

In this case the unspecified individual who could be helped by the extra effort or aid required by either the principle of easy rescue or the principle of heroic sacrifice could reasonably reject the proportionality principle, since that principle will be insufficient to avert the harm that will occur to him. After all, he might point out, the easy rescue principle, while it demands a somewhat higher level of aid than does the principle of proportionality, nevertheless requires no significant sacrifice. It requires only an easy rescue. Similarly, the person who will be obligated to donate in this situation cannot complain or reject the rule, because twenty minutes of button pushing is really no significant burden. Other persons not contributing their fair share is not a good reason here, since my aid is guaranteed to help some individuals regardless of the actions of others. Whatever level of help I provide, it will not be wasted. Thus the reason to reject the easy rescue or heroic rescue principles in non-individuable contexts is not a terribly good reason in this context, since no matter how much or little others give, my giving, at whatever level, will enable at least a few people to avoid the significant harm.

But then why can’t those other individuals who will be harmed reasonably reject all principles except those that require the highest level of aid from each individual, namely the heroic rescue principle? The problem with that principle, however, has already been noted- given how much that principle requires, it can be reasonably rejected by potential donors as being too intrusive of their autonomy.

**On the Diachronic Nature of Help**

One final point needs to be made here. The discussion up until now has assumed that offering aid was a one time affair. The issue of collective action has been posed as though a person will have only one opportunity to help others to avoid significant harm in her lifetime. Unfortunately that is not true. Most people will have many opportunities to provide help to others who are facing significant harm. And if there are enough cases when I meet my obligation and perform easy rescues or give a proportionate amount and that
occurs frequently enough, it may all add up to a heroic rescue, which is a level of help which can be reasonably rejected. I have an obligation to perform an easy rescue and save the drowning toddler, even at the cost of ruining my $100 shoes, and whether or not there are other onlookers who could have acted but didn’t. But suppose I encounter drowning toddlers every day? At some point the sum of what I give up in order to perform all these easy rescues turns the next opportunity for an otherwise easy rescue into a heroic rescue.

Thus the question of how much we ought to give in any particular case of collective action will be partially determined by how much we have done in the past. But how help should we give over the course of a lifetime? My guess is that, given how many collective action situations occur on the planet, meeting one’s obligations for all of these would put a person’s donations well over the level that would be imposed on all of us by the easy rescue principle over the course of her lifetime. Since many of these are individuable situations, our obligation over the course of a lifetime is to donate at a level required by the easy rescue principle.

Bibliography


---

1 Of course in real life we do not frequently encounter drowning children, so given an opportunity to save one we ought to, even if it puts us over the level required by the easy rescue principle. Such an act relieves us of having to perform easy rescues in the future, given that the question of how much we owe must be considered from the perspective of one’s entire life.
According to Coleridge, everyone is born either a Platonist or an Aristotelian. If it is not taken too literally, I would like to make a similar point about Kant and Hegel. There seems to be a tendency among prominent philosophers who have struggled with the problem of global justice to favor one of two quite different methods of arriving at a solution. In response to antinomies in which reasonable arguments are given to support two incompatible conclusions, the Kantian type of solution is two reject both horns of the dilemma and substitute a new framework. On the other hand, Hegel’s dialectical approach sees partial truth on each side and resolves the dilemma by grasping both horns and seeking a synthesis that transcends the one-sidedness of each point of view. Although my rather simplistic contrast between the methods of Kant and Hegel will probably not stand up to scrutiny, I use it merely to illuminate the different styles and strategies of Martha Nussbaum and Amartya Sen in dealing with the contradictory claims in the recent debate about the adequacy of the social contract theory to give an account of global justice. The impasse can be briefly described as follows: John Rawls defines the original social contract as limited to members of a nation whose representatives subsequently contract with those of other nations to form principles of international justice, whereas critics of Rawls have maintained that a just social contract must be global from the beginning and include all human beings. I describe Nussbaum’s strategy as “Kantian” in that she rejects both social contract models as structurally inadequate to provide the basis for a theory of global justice, and she proposes replacing the social contract theory with the “capabilities approach.” Sen’s approach is more “Hegelian” in that he finds a synthesis between the national and global interpretations of the social contract theory in a third alternative which he calls “plural affiliation.” I shall discuss both strategies and argue that in this case the Hegelian approach is
more fruitful than the Kantian. Although Nussbaum’s criticisms of Rawls are strong, they do not have force against social contract theories in general, as Sen’s interpretation demonstrates. I conclude that the social contract theory still has life in it, and Nussbaum’s attempt to bury it is premature.

Nussbaum contends that the social contract theory has serious flaws preventing it from giving an adequate account of global justice. She writes, “We cannot solve the problems of global justice by envisaging international cooperation as a contract for mutual advantage among parties similarly placed in a State of Nature. We can solve them only by thinking of what all human beings require to live a richly human life—a set of basic entitlements for all people—and by developing a conception of the purpose of social cooperation that focuses on fellowship as well as self-interest.” Her objections focus on Rawls, whom she regards as the “greatest modern exponent” of the social contract theory and who she thinks “makes the strongest case for its superiority to other theories.” Her argument is that if she can find serious structural weaknesses in Rawls’ version, then “a fortiori other, less developed or less appealing forms of the contract doctrine are likely to have such problems.”

Nussbaum identifies three features of social contract theories which she regards as “structural defects that make them yield very imperfect results when we apply them to the world stage.”

Her first objection is to Rawls’ idea that “the social contract is made between parties who are roughly equal in power and resources so that no one can dominate the others.” She argues that to be consistent he should say that countries such as India, Bangladesh, and South Africa cannot belong to the contract, because they are “just too poor for the richer nations to gain anything from treating them as equals.” According to Nussbaum, “Rawls has not thought this through; his [lack of clarity] at this point makes The Law of Peoples a very unsatisfactory work.” Other critics have also noticed a serious discrepancy between Rawls’ egalitarian view of domestic justice and his tolerance for enormous inequalities in his theory of international justice. According to Peter Singer,

When I first read [A Theory of Justice], shortly after its publication in 1971, I was astonished that a book with that title, nearly 600 pages long, could utterly fail to discuss the injustice of the extremes of wealth and poverty that exist between different societies. If he accepted that to choose justly people must also be ignorant of their citizenship, his

---

1Martha Nussbaum. (2005). “Beyond the Social Contract,” The Political Philosophy of Cosmopolitanism, ed. Gillian Brock and Harry Brighouse, 197-198. New York: Cambridge University Press. See also Martha Nussbaum. (2005). Frontiers of Justice, 3. Cambridge, MA: Belknap. There is some ambivalence in Nussbaum’s position with respect to the social contract theory. In spite of finding it “wanting,” “deeply flawed,” and in need of “wholesale rethinking,” she writes, “My conclusion is not that we should reject Rawls’ theory or any other contractorian theory, but that we should keep working on alternate theories, which may possibly enhance our understanding of justice and enable us to extend those very theories.”

2Frontiers of Justice, 3.

3Ibid., 4.

4“Beyond the Social Contract,” 197.

5Ibid., 204.
theory would become a forceful argument for improving the prospects of the worst-off people in the world. But in the most influential work on justice written in twentieth-century America, this question never even arises.¹

Rawls’s hypothetical “veil of ignorance,” which prevents people from knowing to which economic, racial, religious, or educational group they would belong, does not extend far enough to conceal knowledge of their citizenship, but he does not explain why nationality is considered to be less accidental to a person’s identity than race or class.

I agree that this is a serious weakness of Rawls’ theory, but an obvious way around it is to give a different definition of ‘equality’ from the “rough equality in power and resources” that Rawls borrows from Hume. The traditional social contract theorists conceived of human equality in the state of nature not as “positive equality” of powers and resources, but rather as what might be called “negative equality” in the sense that everyone is equally vulnerable to harm or violent death. If we imagine a hypothetical “state of nature” among nations in which there are no structures governing international relations, the lack of economic equality among citizens of rich and poor nations is less relevant to the formation of a global social contract than their “negative equality” in the sense of being in relatively equal danger from the threats of nuclear war and environmental crisis. Hobbes and Locke are better guides than Rawls in extending social contract theory to the global sphere.

Nussbaum’s second objection is that contractarian theories derive justice from mutual advantage, “where advantage is defined in familiar economic terms, and income and wealth play a central role in indexing social positions.” She points out the recent “untold influence” on public debate “especially the idea that we ought to expect to profit from cooperation with others,”² and argues for an alternative to the social contract approach in which social cooperation is based on compassion, fellowship, and a love of human dignity rather than mutual advantage.

If our world is to be a decent world in the future, we must acknowledge right now that we are citizens of one interdependent world, held together by mutual fellowship as well as the pursuit of mutual advantage, by compassion as well as self-interest, by a love of human dignity in all peoples, even when there is nothing we have to gain from cooperating with them. Or rather, even when what we have to gain is the biggest thing of all: participation in a just and morally decent world.³

²“Beyond the Social Contract,” 198.
³Ibid., 218; Frontiers of Justice, 324.
Charles Beitz\(^1\) and Thomas Pogge\(^2\) develop a theory of global justice that applies the original position directly to the whole world rather than going through the intermediary of the state. Nussbaum thinks that this approach is “far more appealing” and a “big improvement” over Rawls’s, but she finds it “vague and speculative” about the role of the nation-state and the multinational structures that will govern people’s lives.\(^3\) Nussbaum attributes the advances that this global theory of the social contract makes to “a departure of major proportions from the Rawlsian framework.” She thinks that the one-step global bargain of Beitz and Pogge “requires admitting from the start that the point of the bargain cannot be mutual advantage among equals, but it must be human fellowship, and human respect, in a more expansive sense.”\(^4\)

But surely her sharp contrast between mutual advantage in social contract theory and compassion in the capabilities approach sets up a false dichotomy. Rawls assumes that individuals in the original position will be self-interested, but real-life contracts are often made for the sake of common goals as well as self-interest, which the traditional social contract theorists recognized in their emphasis on the mutual common advantage of leaving the state of nature more than the mutual separate advantage to individuals. Nussbaum’s lack of clarity about the relationship between justice and “mutual advantage” is illustrated in the following passage:

> The capabilities approach denies that the principles of justice have to secure mutual advantage. Even where noncooperation is possible and even habitual (because domination is so easy), justice is good for everyone. Justice is about justice, and justice is one thing that human beings love and pursue.\(^5\)

> The claim that “justice is good for everyone” is inconsistent with the denial that “principles of justice have to secure mutual advantage.” If justice is not necessarily mutually advantageous, then how can it be good for everyone? The tautological assertion that “Justice is about justice” adds nothing to the argument.

The third problem that Nussbaum identifies in the social contract theory is the idea that “contract theories take the nation state as their basic unit,” conceiving of the social contract as a two-stage process in which contracting parties first choose principles for the state, after which representatives of nations join to form a second contract in which they agree upon “the fundamental principles to adjudicate conflicting claims among states.”\(^6\) Nussbaum criticizes this two-step approach, because the members of the second-stage contract may not represent the interests of the citizens, and the interests of groups are emphasized over individuals. She thinks that the

---


\(^3\)“Beyond the Social Contract,” 207; *Frontiers of Justice*, 264ff.

\(^4\)Ibid., 209.

\(^5\)*Frontiers of Justice*, 89.

assumption of the fixity and self-sufficiency of states results in a “very thin and restricted” second-stage contract, consisting only of the principles of just war theory, and there is no “serious consideration of economic redistribution from richer to poorer nations.” She concludes that leaving no room even for “a supranational political/economic structure such as that of the European Union,” let alone the complex interdependencies of the world as a whole, Rawls’s structure “is so far from being true of the world in which we live that it seems most unhelpful.”

According to Nussbaum, Rawls’ later work, *The Law of Peoples*, makes some progress beyond *A Theory of Justice* in resolving these problems by recognizing the “transnational force of human rights,” but she finds his list of rights “thin,” including freedom from slavery, liberty of conscience, and security of ethnic groups from mass murder and genocide, but fewer than half of the rights enumerated in the UN Declaration. She argues that he does not go far enough in his idea of the scope of the duty of assistance to burdened nations, which is limited to helping them develop democratic institutions rather than giving significant material aid. The little progress that Rawls’ *The Law of Peoples* makes over his earlier work is gained, she claims, only by “dramatic departures” from the contractarian approach in the direction of a concept of justice defined in terms of positive outcomes in respect for human rights, bringing him closer to the “capabilities approach” which she endorses as preferable to the social contract theory as the basis for a theory of global justice.

Although Nussbaum argues persuasively that Rawls’ theory does not provide an adequate account of global justice, it is not altogether clear why she thinks that the social contract theory and the capabilities approach are incompatible. Sen gives a non-Rawlsian interpretation of the social contract theory that is consistent with the capabilities theory that he introduced in 1979. Sen has defined the term “capabilities” as the ability to perform certain basic human functions, presenting it as an alternative to the Rawlsian concept of primary goods and the utilitarian focus on levels of satisfaction in theories of development. Nussbaum agrees with Sen that the capabilities approach is a major improvement over economic growth as an indicator of the quality of life in a nation, but she has developed the concept in a way that conflicts with his original purpose. She endorses a list of ten “Central Human Capabilities” as the basic requirements of a life with dignity, and briefly summarized, these include life, bodily health and integrity, the ability to sense, imagine, reason, play, form emotional attachments and affiliations, and to have control over one’s political and material environment. Nussbaum views her list of capabilities as important “both for comparative quality-of-life measurement and for the formulation of basic political principles of the sort that can play a role in fundamental constitutional guarantees, and it is intended to be provisional and open to

---

2 Ibid., 203.
modification in the light of criticism."¹ She has criticized Sen for failing to give substantive content to the idea of capabilities with a list of the minimum opportunities for functioning in a fully human life, but Sen insists on leaving the specific list open for local public discussion and revision, taking account of varying practical purposes and local social conditions as well as the important role of public discussion in leading to a better understanding of particular capabilities. He writes, “I have nothing against the listing of capabilities but must stand up against a grand mausoleum to one fixed and final list of capabilities.”²

The debate between Nussbaum and Sen on whether or not there should be a list of capabilities sheds light on other important differences between their views that are apparent in Sen’s recent works. In “Global Justice: Beyond International Equity,”³ Sen states that a “serious departure” from Rawls will ultimately be needed, but he takes Rawls’ idea of “justice as fairness” as his starting point and uses the concept of the original position, which he regards as a “rich way of characterizing the discipline of reciprocity and within-group universalization.” In answer to the question, “Who are the individuals who gather together in the original position?,” he points out the inadequacies of two possible answers, contrasting what he calls “grand universalism,” the position that all the people in the world, regardless of their nationality and citizenship, constitute the social contract, with “national particularism,” the idea that citizens of each nation form separate contracts within their own countries. He credits “grand universalism” with “an ethical stature that is hard to match in terms of comprehensive coverage and non-sectarian openness,” but he sees difficulties in implementing the rules that such a hypothetical contract would arrive at in the absence of a world-wide institutional base. The perspective of “national particularism” allows only for international, as opposed to more broadly global interactions, and it cannot account for relations, institutions, and obligations among people across borders that are not based on solidarity with national and political units.

Sen argues that we need “a different conception of global justice—one that is neither as unreal as the grand universalism of one comprehensive ‘original position’ across the world, nor as separatist and unifocal as national particularism (supplemented by international relations).” Designating this third conception as “plural affiliation,” a perspective which recognizes that everyone has multiple identities that cannot always be overridden by the idea of citizenship in a nation, he illustrates it with examples of a doctor or a feminist activist, who may have affiliations and commitments to a community that extends beyond national boundaries, as well as nongovernmental organizations such as OXFAM, Amnesty International, and Human Rights Watch, which

explicitly focus on global issues rather than on relations among nations. ¹ Sen challenges the ideas that the exercise of social contracting must be applied only once and that it requires an elaborate system detailing the “stage-by-stage emergence of basic structures, legislation, and administration.” He employs the device of the original position in a “less grand, less unique and less fully structured forms” in order to provide “insights and inspirations for different group identities and affiliations” without giving priority to a single nationalistic contract. Sen’s move from Rawls’ starting point to a radically different conclusion turns the social contract theory upside down and, in my view, places it on its feet. He shifts the focus of the theory to the real multiple agreements that people make every day, as opposed to a single hypothetical contract, which has for centuries been used as a rationalization for either obedience to or rebellion against a national authority.

In “Open and Closed Impartiality,” Sen further departs from Rawls by distinguishing different ways of interpreting the conception of the original position in social contract theory. Rawls’ method imagines the exercise of impartial assessment to be confined to a fixed group, and no outsider is a party to the contract, nor are any perspectives other than those of the focal group included. In the case of open impartiality, on the other hand, judgments from outside the focal group must be invoked in arriving at an impartial assessment. Sen uses Adam Smith’s concept of “the impartial spectator” in order to illustrate a procedure for achieving impartiality that is “open and broad, rather than closed and restricted.”² Smith points out that we can never survey and form judgments about our sentiments and motives unless we try “to view them as at a certain distance from us. But we can do this in no other way than by endeavouring to view them with the eyes of other people, or as other people are likely to view them.” According to Sen, “This way of invoking impartiality, including ‘real spectators’ (distinct from social contractors), has a very different reach from the impartiality in the closed form of contractarian reasoning.” The closed model of impartiality is geared to address sources of bias within a particular group, but it does not offer a procedure for overcoming prejudices that may be held by the group as a whole by means of an open scrutiny of local and parochial values and practices, e.g. stoning adulterous women, selective abortion of female fetuses, or capital punishment.³

In his recent article in The Journal of Philosophy, “What Do We Want from a Theory of Justice?,” Sen distinguishes between the what he calls the “transcendental” approach, which focuses on identifying criteria for a perfectly just society, and the “comparative” approach, which concentrates on ranking alternative societal arrangements as more or less just. Rawls and many others follow the first method, but Sen favors the second, arguing that a transcendental theory, which is like a “grand revolutionary’s complete handbook,” is neither a necessary nor a sufficient condition for making

²Ibid., 451.
³Ibid., 459.
comparative judgments in the effort to advance justice or to reduce the many injustices of the world. Sen criticizes Rawls’ restriction of the original position to members of a nation on the grounds that people in other countries may be affected by national policies, such as global warming or the “war on terror,” and their experience and disinterested perspective may useful, especially in challenging parochial prejudices that are widely shared. He agrees with Rawls’ emphasis on the importance of open and interactive public reasoning in theorizing about justice, but he insists that the deliberations not be limited by either a perfectionist ideal or by national boundaries. He concludes, “We have good reason to abstain from concentrating so fully on the program of identifying the totalist—and possibly parochial—demands of transcendental, contractarian justice. We have to move the theory of justice out of that little corner.”¹ In his recent book entitled Identity and Violence Sen ties together a number of themes from his previous works by exploring the ways in which the assumption of a single identity can lead to violent confrontations. He points out that Samuel Huntington’s influential idea that the “clash of civilizations” will be the main source of future conflicts presupposes the questionable view that the primary way of categorizing people should be in terms of their “civilizational identity,” however that vague notion may be defined. The upshot of this kind of unthinking analysis, he argues, is to put people into “little boxes,” which “miniaturizes” them and denies their freedom to choose their identities.²

We can now sum up how the positions of Nussbaum and Sen diverge. Nussbaum objects to the idea that global justice can result from a social contract among roughly equal nations for the sake of mutual advantage, and she points out that Rawls fails to address glaring injustices in the economic inequalities among nations as well as the treatment of disabled persons and animals. She supports a capabilities approach to a theory of global justice, which emphasizes obligations to non-equals and is motivated by compassion rather than self-interest. I began by describing Nussbaum’s method as broadly-speaking Kantian in that she resolves the antinomy between the two main schools of contractarian thinking about global justice by undermining both of them. There are strong Kantian elements in the substance of her analysis as well, e.g. her argument that the motive behind social cooperation should not be self-interest and her commitment to a universal list of categories of human capabilities as the necessary condition for a theory of global justice. In Sen’s terms we could say that Nussbaum’s approach illustrates the “transcendental” model of defining justice in terms of ideal conditions, as opposed to the “comparative” model of piecemeal reform.

In contrast, Sen accepts the central principle of the social contract theory that the rational deliberation of the people is the sole source of political legitimacy, but he denies that the social contract is a singular and one-time agreement either among nations or in a global contract of all people. While accepting Rawls¹ idea that justice is fairness, he doubts that fairness is likely to

Global Justice after Rawls:
Kantian and Hegelian Strategies

be achieved in a fixed group without the perspective of “impartial spectators.” He develops a model that is pluralistic in recognizing that individuals who have multiple identities and affiliations can enter into a variety of social contracts, both within and across national borders, and it is open in its demand for the kind of impartiality in which the perspectives of those outside the original group are included. Sen advocates thinking about justice in comparative rather than transcendental terms, which attempts to take steps, however partial and imperfect, towards reducing specific injustices. He sees no incompatibility between non-Rawlsian contract theory and the capabilities approach. He disagrees with Nussbaum about the need for a definitive list of basic capabilities, arguing that this should be left open for local public discussion, consensus, and revision, a position that is consistent with his model of plural, open, and continually renegotiated social contracts. I interpret Sen’s position as broadly-speaking Hegelian, not only in its method of synthesizing a number of traditions and perspectives, global justice through the interaction of multiple points of view based upon public discussion and local consensus. Of course, there is no hint in Sen’s view of the but also in its emphasis on the “dialectical” development of a theoretical perspective on Hegelian idea of Absolute Knowledge, and he rules out the possibility and desirability of such an ideal perspective in rejecting the transcendental approach. Sen points out the importance of intellectual fairness in global history in order to achieve a more complete understanding of our human past as well as overcome the false sense of Western superiority that has contributed to artificial and simplistic confrontations between Western and anti-Western identities.

Nussbaum’s approach to the problem of global justice borrows from the Aristotelian tradition and Grotius’ natural law theory as well as Kant, and it has been rightly praised for making a valuable contribution to a liberal reading of value pluralism.¹ Her theory of human capabilities draws the strengths of elegance, universality, and impartiality from the classical view of rationality, but it shares the weaknesses as well as the strengths of this traditional ideal. The assumption that the purpose of a philosophical theory of justice is to provide an account of universal and unchanging aspects of human nature that justify basic moral principles is vulnerable to Sen’s criticism that it exaggerates the role of pure theory in ethics. If the alleged legitimacy of universal and invariant principles of human nature and morality can be shown to derive from imposing the particular philosophical and political values of a dominant group on humanity as a whole, there is the danger of erecting what Sen calls a “grand mausoleum” which would block the way of further inquiry and public discussion.

Sen’s Hegelian approach, by remaining closer to actual human practices and giving up the idea of a single social contract, challenges the traditional assumption that the goal of a philosophy of global justice is to seek a unitary theoretical model. While avoiding the criticisms that may be made to

Nussbaum’s approach, Sen’s method may be viewed from the perspective of the classical and Kantian traditions as vulnerable to different kinds of problems. It could be argued that Sen’s denial that there are universal criteria for moral judgment prevents him from producing a rational theory of justice at all. Some might go so far as to accuse Sen of being a postmodernist, which according to Rorty is “the worst thing one intellectual could call another.” Following this line of argument, we now seem to face an even deeper and more serious dilemma than the impasse between particular and global ways of interpreting the social contract with which we began. Nussbaum’s alternative means replacing the social contract theory with a version of Sen’s capabilities model that he rejects as too fixed and final. On the other hand, it could be argued that Sen’s pluralistic model of the social contract theory falls into relativism and undermines the ideal of rationality and the traditional role of philosophy.

It seems to me, however, that the appearance of an inevitable choice between the devil of the “grand mausoleum” and the deep blue sea of postmodernism is a false dilemma. Sen steers skillfully between the horns by holding on to the idea of public discussion aimed at agreement, which is the essential concept of the social contract theory, while letting go of the demand for universality. He points out that in a variety of circumstances individual communities will have different and changing priorities about the importance of specific capabilities to solve problems in such areas as food, shelter, education, and communication. His concepts of open impartiality and the impartial spectator give models for resolving conflicts through open public debate, avoiding both narrow parochialism and the imposition of universal principles from outside. While eschewing the utopian ideal of universal agreement, his model escapes relativism by emphasizing the goal of impartiality, defined as openness to the points of view of outsiders and impartial spectators. Sen pluralizes and historicizes the traditional social contract theory by broadening it to include multiple identities and by basing it on actual human social practices rather than hypothetical principles. This way of thinking is less parochial and nationalistic and more open to renegotiation than Rawls’ approach, which makes it more relevant for resolving the problems of global justice.

Nussbaum claims that her critique of the social contract theory removes an obstacle to the realization of global justice by establishing “that a particular picture of who we are and what a political society is has for some time imprisoned us, preventing us from imagining other ways in which people might

---

1Richard Rorty. (2002). “To the Sunlit Uplands,” Review of Truth and Truthfulness: An Essay in Genealogy by Bernard Williams. London Review of Books 24(21). In developing this hypothetical line of argumentation, it is far from my intention to imply that I think Sen is a postmodernist, whatever that might mean. The term “postmodern” has lost whatever distinct sense it may once have had through indiscriminate overuse, and is now mainly a term of abuse, like the word “bastard,” which once meant the child of unmarried parents and now means anyone who is disliked. Similarly, the label “postmodern” is now used primarily in ad hominem rhetoric by those who believe in the existence of universal principles, objective moral values, the correspondence theory of truth, methodological monism, and other traditional philosophical views against anyone who disagrees with them.
get together and decide how to live together.”¹ She has a good point but aims it at the wrong target by rejecting the social contract theory entirely because of weaknesses in Rawls’ interpretation. The central insight of the traditional social contract theory is the idea that everyone can see by the “light of reason” the advantages of controlling violence through collective agreement rather than force. Today we would substitute for the rather nebulous expression, “light of reason,” the concept of “open public discussion,” but the idea has not lost any of its persuasiveness or importance. Sen has shown how a pluralist and historicist interpretation of the social contract theory can rescue it from the abstractness of the Kantian tradition, thus revitalizing the role that it has played for centuries as one of the main philosophical flagships in the fight for global justice.²

¹Frontiers of Justice, 414.
²I am grateful to Martha Nussbaum for her encouragement and for her detailed comments on an earlier version of this paper.
Competition and Reasons of Autonomy: A Rejection of Korsgaard’s Theory of Reasons

Heather Salazar, Spring Hill College, USA

On Korsgaard’s view, reasons are generated by the reflective endorsement of desires by agents. However, the reasons generated by a particular agent do not merely provide reasons for that agent. Since all reasons are agent-neutral on her account, every desire endorsed by a desiring agent also produces complimentary reasons in other all other rational agents to help the desiring agent achieve the object of her endorsed desire.

Korsgaard’s account of the source and nature of reasons can be seen as a compromise between the two traditional theories of the source and nature of reasons: realism and subjectivism. Because reasons are agent-neutral, they function like objective values for all of humanity (like realism), yet unlike realism, these reasons depend on our mental states and not on values in the world (as subjectivism asserts).

It is vital to note that Korsgaard argues extensively for the claim that all reasons are agent-neutral as it is a crucial premise in her argument to show that constructivist accounts can produce moral reasons. If reasons are agent-relative, she thinks that constructivism cannot succeed in explaining how we can be obligated to one another.

1Korsgaard says: ‘[R]easons springing from ambitions are agent-neutral. But they spring from our respect for one another, rather than from our respect for one another’s ends’ (‘The Reasons We Can Share,’ 290).

2See Lecture 3 of Korsgaard’s The Sources of Normativity for her explanation of the origin of reasons.

3In Part 2 of my dissertation, Meaning, Reasons, and Endorsement, I argue that Korsgaard’s theory cannot accommodate agent-relative reasons; I show that admitting agent-relative reasons into her theory would render her account deeply inconsistent.
Reasons of Autonomy: The Natural Reading

Reasons of autonomy, which involve the interests and ambitions of particular agents, present a challenge for Korsgaard’s view, as they are naturally seen as being agent-relative, providing reasons only for the specific agents who possess the ambitions. For example, my friend Holly’s ambition to buy and display the mugs that commemorate each of the 50 states in the Starbucks’ collector coffee mug line appears to give only her a reason to seek and buy these mugs; it doesn’t give me a reason to help her. So a proponent of Korsgaard’s theory carries the burden of explaining how these reasons can be plausibly interpreted as agent-neutral. Realists about reasons, like Nagel, can explain why reasons of autonomy should be interpreted as agent-relative by claiming that the objects or activities involved in the ambitions have only agent-relative value. Collecting mugs for each of the 50 states is not something good-in-itself, so the only good that it can have is derived from our attitudes towards it. So the realist can conclude that ambitions are the sorts of things that give only the person who has them reasons to achieve them and reasons of autonomy are thus justified as being agent-relative.

Reasons of Autonomy as Agent-Neutral: A First Attempt

Korsgaard’s account of reasons, however, necessitates that all reasons be interpreted as agent-neutral, where all reasons produce normative demands on all agents. Reasons of autonomy, therefore, which are naturally read as agent-relative, deserve special attention on her account. I will examine two ways of seeing how reasons of autonomy could generate agent-neutral requirements, and I will begin with what I call ‘The Intuitive Account.’

The intuitive way of understanding reasons of autonomy as agent-neutral begins with an account for the generation of reasons. It seems that reasons of autonomy, which involve the interests and ambitions of particular agents, originate with the desires of agents to achieve something (as in the agent-relative ‘natural’ reading). As Korsgaard’s account of the genesis of all reasons also grounds reasons in the desires of agents, I will begin by sketching a possible Korsgaardian account of the production of reasons of autonomy, using Korsgaard’s example in “The Reasons We Can Share” of (I’m assuming) her ambition to write a book on Kant:

**The Intuitive Account:**
Step 1: I want to write a book on Kant good enough that it will be required reading (first-personal desire).
Step 2: I endorse that desire upon reflection.
Step 3: This gives me a reason to write a book on Kant good enough that it will be required reading (reason).
Step 4: This reason, in turn, gives everyone a complimentary reason to help me to write a book on Kant good enough that it will be required reading (complimentary reason).
The natural reading of the reason in Step 3 is to say that it is agent-relative, providing only a reason for Korsgaard to write a book on Kant. It seems like it is agent-relative because it arises from a desire for a specific agent to pursue a special interest of hers, and the resulting reason is in reference to her alone. This, in turn, seems to give only her a reason to pursue her own special interest. So ‘the natural reading’ of the reasons of autonomy appears to be reinforced even on Korsgaard’s constructivist account of the creation of reasons from desires.

However, according to Korsgaard, all reasons are actually agent-neutral, so the natural reading must be wrong. The reason in Step 3 that looks like it is agent-relative is actually agent-neutral, so the reason in Step 3 must be generalized to create complimentary reasons for everyone to act so as to help Korsgaard with her ambition. We might reasonably think that step 3 then entails Step 4. But Step 4 doesn’t seem right. Surely everyone does not have a reason to help Korsgaard with her personal project to write a book on Kant.

A Problem for Korsgaard

The move from Step 3 to Step 4, from what seems to be only a reason for a particular person to pursue the object of her ambition to a reason for everyone to help her to achieve that object, is suspect. It does not seem that all people do in fact have a reason to help every individual to achieve her personal ambitions. Furthermore, this view leads to an unintuitively large number of reasons to help people to do everything that they desire to do upon reflection since we have millions of desires that we endorse throughout each of our lives. With over six billion people in the world, if we each have two hundred desires per day that we endorse, then we each have over 1.2 trillion reasons each day, most of which are in reference to helping others with everything they have reason to pursue in their lives. But this seems plainly counterintuitive. It seems like we only recognize that we have a reason to help someone with something if she needs help, and not simply because she wants it, whereas we recognize the force of all of our own desires upon ourselves, great and small, basic and non-basic, necessitating help and not necessitating help.

Two types of response are then available: One accepts the intuitions as providing a framework of limitations for which the theory should account. The other denies that the intuitions reflect reality and dissolves them by explaining their purely psychological basis. A Korsgaardian might take either tack, but Korsgaard uses the first, developing an alternative account of ambitions and reasons of autonomy. After accepting these intuitions as legitimate constraints for her theory, there are then two ways in which she might accommodate them: (a) she could reject that ‘reasons of autonomy’ provide anyone with reasons, and thus revise her account and eliminate reasons of autonomy, or (b) she could dispute the way that ambitions and reasons of autonomy have been interpreted. Korsgaard argues for the latter, maintaining that an improved account of ambitions reveals that reasons of autonomy do not have the
‘intuitive’ structure that is displayed in the example given in the previous section.

**Reasons of Autonomy as Agent-Neutral: Korsgaard’s Alternative**

In ‘The Reasons We Can Share,’ Korsgaard contrasts two accounts of what it is to have an ambition, endorsing one which grounds an alternative explanation of how and in what way reasons of autonomy are agent-neutral. This new way of viewing the agent-neutrality of reasons does not leave the account open to the difficulties discussed in the last section since it does not legitimize our asking others to help us achieve the objects of our endorsed desires simply because we want them.

Korsgaard claims that Nagel’s account of the mental states involved in an ambition is inaccurate. Nagel’s account of the ambitions of agents begins with an agent-relative end:

1. I want my book to be required reading (where what is an agent-relative end);
2. therefore: I shall write a good book (as a means to that end)\(^1\)

Whereas her account begins with an agent-neutral end:

1. Someone should write a book on Kant good enough that it will be required reading (where that is an agent-neutral end);
2. I want to be that someone (agent-relative motive).\(^2\)

Korsgaard defends her interpretation of ambitions by saying that ‘to have a personal project or ambition is not to desire a special object which you think is good for you subjectively, but rather to want to stand in a special relationship to something you think is good objectively.’\(^3\) Korsgaard admits that her account of ambitions seems to presuppose that there is something that is good objectively (the agent-neutral end), yet her account of reasons is opposed to this assumption since all reasons are created from endorsed desires and are not formed from the recognition of agent-neutral values.\(^4\) She does not explore what the combination of her theory of the origin of reasons and this account of ambitions would look like. However, when united, I think that it is clear that these two elements create a tension within the account, rendering it implausible in three ways.

Considering Korsgaard’s new interpretation for ambitions and the account she gives of the genesis of reasons (in which reasons always begin with a desire), it appears that two desires are functioning here:

\(^1\)‘The Reasons We Can Share,’ 287 (direct quote).
\(^2\)‘The Reasons We Can Share,’ 288 (direct quote).
\(^3\)‘The Reasons We Can Share,’ 288.
\(^4\)See ‘The Reasons We Can Share,’ 289.
Desire 1 (Agent-Neutral Desire): that someone should write a book on Kant good enough that it will be required reading.
Desire 2 (Agent-Relative Motivating Desire): that I be that someone (who writes a book on Kant good enough that it will be required reading).

Because reasons for Korsgaard are agent-neutral, the first desire evidently gives rise to a reason for everyone while the second gives rise to a reason for no one; it remains only a desire, providing motivation for the person who possesses it. I call the first type of desire an ‘agent-neutral desire’ to capture the idea that these desires are general and are directed towards an agent-neutral end. Likewise, I call the second type of desire an ‘agent-relative motivating desire,’ since it merely provides motivational force (and has no reason-giving force) for the person who has it.

So in the example that Korsgaard gave regarding her ambition to write a book on Kant, everyone has an agent-neutral reason to see to it that someone writes a book on Kant good enough that it will become required reading. Although Korsgaard is motivated to see to it that the goal is accomplished by writing the book herself, she might also comply with the reason by helping someone else to write it. All people share the same reason to see that the book is written, and people can be motivated to do this in different ways. These differing agent-relative motivations do not mean that some have reasons that others do not since these motivations are not themselves reasons.

Combining a constructivist account of the genesis of reasons and Korsgaard’s interpretation of ambitions reveals an alternative model for the nature and origin of reasons of autonomy. Once again, I use Korsgaard’s example:

**Korsgaard’s Account:**
Step 1’: I desire that someone write a book on Kant good enough that it will become required reading (agent-neutral desire).
Step 2’: I endorse that desire upon reflection.
Step 3’: This gives me a reason to see to it that someone writes a book on Kant good enough that it will become required reading (agent-neutral reason).
Step 4’: I desire that I be that someone (motivating desire).

The reason that is represented in step 3’ is supposed to be an agent-neutral reason which is somehow valuable apart from the particular desires of any one agent. The content of the reason in Step 3’ reflects this change to the extent that it does not restrict the writing of a book on Kant to Korsgaard alone (who is the ‘me’ in this example), but rather it refers to an unspecified ‘someone.’ However, this reason still appears to be the sole property of the endorsing agent. Step 3’ reads: This gives me a reason to see to it that someone writes a book on Kant…. Korsgaard is the one who has the reason to see to it that someone writes the book. Furthermore, a merely superficial repair of 3’ to include all others does not fully apprehend the problem, which begins in the
first and second steps. The extension of the reason that is produced to everyone through Korsgaard’s endorsement of her desire does not appear to be licensed.

This step-by-step rendering also fails to capture Korsgaard’s complete intent, since according to her interpretation of reasons of autonomy ‘[you] want to stand in a special relationship to something you think is good objectively.’¹ Her discussion illuminates an important point, namely, that we can all see that there is something good about many of the objects of people’s desires, but we do not see all desires as being good. She indicates that this explanation of ambitions appears to suppose realism, where values exist independently of our desires.² She does not attempt to explain how an agent-neutral end might exist on a constructivist theory, where values are created from desires, so I will briefly explain how we might understand them on her account.

On Korsgaard’s account, we might explain the goodness of the agent-neutral end as springing from an agreement among us, as members of humanity, that we can see value in pursuing certain objects of our desires. This agreement then might be used to validate those desires that each of us has to pursue them. The fact that we collectively see value in writing a good book on Kant makes it acceptable for Korsgaard to want to write one. If they were merely desires that we had, whose objects contained no apparent good (or good that is understandable intersubjectively, apart from an individual’s standpoint) and the good of it were just to satisfy our own longings to excel, then Korsgaard says that the desire would be criticizable as being an object of vanity and containing no reasoned force whatsoever. For example, she says of her desire to have a statue of herself erected on campus that it is ‘plausible’ to say that it is merely vanity and thus should not be accepted as a reason.³

Steps 1’, 2’, and 3’ of the alternative account are revised below to better reflect these considerations:

Korsgaard’s Account:
Step 1’': I desire that someone write a book on Kant good enough that it will become required reading and so do others (agent-neutral desire).
Step 2’': I endorse that desire upon reflection and so do others.
Step 3’': This gives me and everyone else a reason to see to it that someone writes a book on Kant good enough that it will become required reading (agent-neutral reason).
Step 4’': I desire that I be that someone (motivating desire).

On this account, the reasons that people have are agent-neutral and generated from desires, but their source is in desires that are intersubjectively

¹ ‘The Reasons We Can Share,’ 288 (my emphasis).
² See ‘The Reasons We Can Share,’ 289.
³ See ‘The Reasons We Can Share,’ 289. Although it is tempting to explain the agent-neutral end this way, it is inconsistent with her discussion of reasons in response to Cohen in the last few pages of The Sources of Normativity, 257-258. It is also incompatible with the foundations of her account as fully constructivist. I argue in Part 2 of my dissertation, Meaning, Reasons, and Endorsement, that her account in this section of Sources is incompatible with several very central aspects of her general theory, and also that it undermines her motivation to develop a constructivist account altogether since it results instead in a version of subjectivism.
endorsed, thus validating them as appropriate agent-neutral reasons. The agent-neutral reasons that emerge on Step 3 provide the reason-giving component in people’s ambitions, while the personal component of reasons of autonomy is analyzed here as an agent-relative motivating desire, which should not be confused itself as a being reason.\footnote{See ‘The Reasons We Can Share,’ 288.}

Although both accounts that I have considered for understanding reasons of autonomy render them agent-neutral, the content of the reasons on each of these accounts is different. On the intuitive model that I gave in Section 3, the agent-neutralized reason was a reason for everyone to see to it that I write a book on Kant good enough that it will become required reading. On the alternative model based on Korsgaard’s re-interpretation of ambitions, the reason is a reason for everyone to see to it that someone writes a book on Kant good enough that it will be required reading. However, in order for this revised account to work, two specific events must occur: (1) The agent (Korsgaard, in the example we have been discussing) and others must endorse the agent-neutral desire (as seen in step 2”) transforming it into a reason (thus making it an agent-neutral reason as seen in Step 3”), but (2) the agent must refrain from endorsing the motivational desire (in step 4”), or it must be rendered impossible to endorse, so that it does not become a reason.

The revised structure of reasons of autonomy appears to be a better alternative to the intuitive account that I gave in Section 3 since it sidesteps the problem of there being too many reasons of autonomy. First, we have perhaps not one hundred reasons to help one hundred different people to write a book on Kant good enough that it will be required reading but instead one reason to help see that a great book on Kant is written. Second, the revised construction of reasons of autonomy gives us a way to escape regarding as reasons the ambitions of those who desire objects merely for the sake of vanity since we do not see value in things that are merely projects aimed to further someone’s vanity. There are thus far fewer reasons of autonomy than we might have thought would arise on an account like Korsgaard’s, where all reasons are agent-neutral, yet created from desires.

\textbf{Objections to Korsgaard’s Interpretation}

Although Korsgaard’s alternative interpretation of ambitions and reasons of autonomy appears to resolve the initial problem that was posed, namely, that there are too many reasons and that we should not be required to help people with personal projects that appear to us to be worthless, there are several reasons why I think that her revision to the structure of reasons of autonomy fails.

First of all, the revised structure does not seem to capture the way that we actually reason when we have a personal project or ambition. It does not seem to me that I think that something would be good for someone to accomplish, and that I am simply motivated to do it myself. I do not think: It would be great
if someone collected stamps! I know, I want to be that person! And I certainly do not think: It would be great if someone wrote a book on Kant good enough that it will be required reading…. I want to be that someone! One key feature that reveals that this is not an accurate description of our reasoning process is that in most ambitions, someone has already accomplished the stated goal. Someone has already written a book on Kant so good that it is required reading. (There are Henry Allison’s and Paul Guyer’s books just to name two.) Someone has already taken up the hobby of stamp-collecting, and if having a great stamp collection is a good thing that someone should do well, then there are many people who have an edge on that over me. If it’s just a good thing that someone do it, and someone has already done it, then what reason is left for me to try to do it too?

We cannot claim as a defense against this charge that these ambitions are good pursuits for everyone. That is why we call these sorts of reasons ‘reasons of autonomy’ and ‘personal projects and ambitions.’ Furthermore, it would be theoretically unappealing and dubious if we changed the ‘someone’ in these reasons to ‘some people, whose number is unspecified but which includes at least one more person,’ just in an effort to accommodate myself and anyone else who has the ambition. The revised account of the structure of reasons of autonomy does not therefore give me any reason to do anything that has been done before.

Second, in order for this picture to work under Korsgaard’s account of the source of reasons, it is necessary for the agent-neutral desire to be endorsed and for the agent-relative motivating desire not to be endorsed, as I indicated in the previous section. But this is implausible. Surely it seems possible for someone to endorse these sorts of personal desires, and it appears that most if not all of us do just that nearly every day of our lives. Furthermore, it seems that if the motivating desire motivates the agent, the agent must at least implicitly approve of her desire. What then prevents it from becoming a reason on her account? Either Korsgaard’s theory that reasons originate in the desires of agents is radically mistaken or her interpretation of reasons of autonomy as springing from agent-neutral ends is incorrect.

The third reason why I think that this revised account of reasons of autonomy should not be accepted is that it cannot make sense of the idea of competition, which is not only a very significant aspect of our lives and of why we do things, but also one of the central features in explaining reasons of autonomy, which very often include competitive aspects.

Many, if not most, ambitions have a competitive element. I don’t just want to write a good book on Kant, I want to write the best. I don’t just want to produce good art, I want it to be recognized and distinguished. Many other reasons of ambition are even more explicitly competitive and would be not only weakened but actually nullified were we to accept Korsgaard’s account of the structure of reasons of autonomy. For example, every reason that participants in competitive games thought they had to win would be discarded. Beyond this are the intrinsically competitive activities that comprise some of our most important life activities. There is competition for employment and
distinction, for renting and buying housing, in dating and courting, and in practically every aspect of business endeavors.¹

Korsgaard’s revised structure for reasons of autonomy would have to phrase the agent-neutral aspects of such ambitions in a way something like this:

1. Someone should win this competition, to the exclusion of all others involved.
2. Someone should be better than or beat the last person who did this.

But there is a problem with these kinds of descriptions of competitions. If the reason that sets the normative goal is that someone should win the competition, whatever it may be, to the exclusion of the others, and my specific goal of winning is not a reason, but simply a motive, then it seems that the goal of the reason would be best accomplished by having all of the competitors but one relinquish their motives to be the one who wins the competition. All rational competitors, under this model, would have a reason to try to help the one who is most likely to succeed to win. This, furthermore, would have an ironic impact on competitive activities, making them cooperative but without the promise of helping all of the parties involved, and thus appearing to change the very nature of competitive activities. It would, in addition, have the taxing psychological consequence of rationally requiring all but one competitor to frustrate what they would naturally call their own ambitions (which under Korsgaard’s description are their motives), a requirement that is plainly too demanding and works in a way contrary to what this revisionist solution was supposed to achieve.

Korsgaard herself recognizes that her account eliminates competitive ambitions, but fails to understand the unintuitive consequences. She says:

If I took it seriously that my desire that I should be the one to write the book was a reason for action, then I would have a reason to prevent one of the other Kant scholars from writing her book. But in fact, neither I nor anybody else thinks I have a reason to do this, even if in competitive moments I am tempted to feel it. This is not an expression of ambition, but rather a very familiar perversion of it.²

Korsgaard’s account eliminates all competitive ambitions, including those that we find necessary and rewarding.

Korsgaard’s new approach to reasons of autonomy looks at first to be a promising alternative to the other two approaches. By reading reasons of autonomy as agent-neutral instead of as agent-relative (as on the natural reading I presented), she is able to argue for a constructivist theory of morality. By explaining the agent-neutral content of reasons of autonomy as springing from agent-neutral ends, rather than from agent-relative reasons which are generalized (as on the intuitive account I sketched), the account escapes

¹Jennifer Scott brought these latter examples to my attention.
²‘The Reasons We Can Share,’ 288.
difficulties with demanding that people help everyone with their ambitions. However, the revised structure does not capture the way that people actually reason, it produces conflicts and inconsistencies within Korsgaard’s account, and it cannot account for the widespread and commonsensical notion of reasons that involve competition. Korsgaard’s alternative account therefore fares even worse than the intuitive account that I gave on which each person has a reason to help everyone else with the objects of their ambitions simply because they want their ambitions to be satisfied. Since this account is also extremely implausible, as I argued earlier, two options remain: (1) we might choose to modify Korsgaard’s theory substantially in order to include agent-relative reasons, or (2) we might choose to discard her theory altogether. As for the first alternative, as I explained in Section 2, Korsgaard’s central argument for moral obligations relies on the agent-neutrality of reasons and would therefore fail. If she is right, and the agent-neutrality of reasons is necessary for the creation of morality, then constructivism cannot present a consistent and plausible alternative to the traditional theories of ethics. We are then once again left with subjectivism and realism.

---

1I argue in Part 2 of my dissertation, *Meaning, Reasons, and Endorsement*, that the admission of agent-relative reasons on a Korsgaardian account produces a severely inconsistent theory of reasons.
At first glance, both the Platonic and Buddhist traditions seem to issue in similar and unambiguous condemnations of the role of desire in the good life. On the one hand, Plato reviles desire as a ‘compelling, goading power’ which represents a serious impediment to reason. The source of all battles, wars, and revolutions, it is ‘a great deal of nonsense with the result that we literally never get an opportunity to think at all about anything.’ On the other hand, the Buddhist Dhammapada describes desire as ‘fierce craving, full of poison.’ Those who experience desire ‘run about like a hunted hare…fast bound in its fetters, they undergo suffering for a long time, again and again.’ Thus, on both accounts, the best life is one on which desire does not represent the impetus for choices and actions.

These condemnations of desire seem to be in direct conflict with contemporary action theory. On many current philosophical accounts, desire – or, at least ‘wanting’ - is a necessary condition of intentional actions. Thus, for example, Robert Audi argues that wanting is not only the ‘most basic...
motivational notion,’ but is also ‘fundamental in the concept of action.’ Even further, he claims that these considerations ‘make it quite plausible to hold that intending entails wanting.’ Davidson’s classic ‘Actions, Reasons, and Causes’ takes a similar position. Arguing that intentional actions follow from an agent’s ‘having some sort of pro attitude toward actions of a certain kind’ and believing ‘that his action of that kind,’ he concludes that desire is necessary for any sort of agency. Finally, Wayne Davis concludes: ‘Without the principle that our actions are caused by our beliefs and desires, it would be hard to understand how it is possible for us to control what we do, or for our actions to be up to us.’

Thus, this paper considers the Platonic and Buddhist accounts, on which desire is ruled out as an appropriate source of motivation, in light of contemporary action theory, on which it is required for intentional actions. While it is clear that each of the classic accounts denounces the role of desire in human action, important questions remain: What conception of ‘desire’ underlies the accounts given in these texts? Are these conceptions commensurable? Are they tenable? Most importantly: given these conceptions, are the classic rejections of desire appropriate? What implications do they have for our understanding of our intentional actions in the good life?

Plato’s ‘Mad Desires’

Plato’s rejection of the value of desire in the good life, of course, reflects his account of desire, on which it is the source of irrational motivation toward bodily gratification. In the Republic, desire is described as that irrational part of the soul which aims at physical pleasures – that ‘with which it loves, hungers, thirsts, and feels the flutter and titillation of other desires…[it is] the companion of various repletions and pleasures.’ As such, to be motivated by desire is to behave in a way which is wholly unreasoning - to follow only impulses and appetites toward frivolous pleasures, and make no discriminations even between the relative merits of the objects of its desire: ‘…thirst will never be desire of anything else than that of which it is its nature to be, mere drink, and so hunger or food.’

5Republic, trans. Paul Shorey, in Plato’s Collected Dialogues, ed. Edith Hamilton and Huntington Cairns Princeton, NJ: Princeton University Press, 1961), 439d. There is an interesting distinction between the account given in the Phaedo, and the account given in the Republic and the Phaedrus. On the one hand, in the Phaedo, Plato identifies the desires with the ‘human, mortal, multiform, unintelligible, dissoluble, and never self-consistent’ stuff of the body, which is treated as utterly distinct from the ‘divine, immortal, intelligible, uniform, indissoluble, and ever self-consistent and invariable,’ stuff of the soul. The body – and its desires – are described as ‘infesting’ and ‘distracting’ the pure, reasoning soul. On the other hand, in the Phaedo, Plato, he treats desire as an actual part of the soul, which is closely linked to the body.
6Ibid., 437d-e.
Further, these irrational urges represent a dangerous distraction from the real work of pursuing wisdom. Thus, on Plato’s account, the philosopher recognizes ‘the soul can best reflect when it is free of all distractions such as hearing or sight or pain or pleasure of any kind – that is, when it avoid[s] all physical contacts and associations as much as it can, in its search for reality.’ In the best life, desire must be controlled, lest it ‘penetrate’ or ‘infect’ the reason with the wrong objects – and ‘drag us irrationally toward pleasure.’

As if that isn’t bad enough, Plato also asserts that desire also has exceptional power to motivate irrational and wrong actions. Desire is described as a motive-source which, once accepted, is both incredibly compelling and rapidly escalating. Plato’s description of the effects of the intensification of desire is not pretty: ‘[the appetites and pleasures] inject this new master of a man's soul with the itch of desire. Stung to frenzy, it chooses madness for its bodyguard and goes berserk. It there are any worthy opinions or desires left in the man - if he is still capable of shame - he will be bled and purged of all these. He will be emptied of sound mindedness, and madness will be imported into his soul.’ At the furthest extreme, desire-driven tyrants move from dissipated living - ‘feasts, reveling, prostitutes, orgies, and the like’ to the commission of paradigmatic Greek evils such as theft, kidnapping, and parental abuse. In the final instance, Socrates asserts that the desire-driven life is fully seven hundred twenty-nine times as unhappy as that of the individual whose desire is controlled by reason!

Thus, it is clear that at least part of what lies behind Plato’s rejection of desire is his view of it as essentially irrational – as a brute physical drive, not consequent on any sort of reasoning, which identifies the wrong objects to pursue, and so leads us away from the pursuit of wisdom. In addition, he fears its strength – dangerously prone to intensification, it not only serves as the irrational motivation for inappropriate actions, it does so with sufficient strength to overwhelm reason.

Interestingly, this is essentially a development of an earlier Greek view identified by Prodikos: that as ‘epithumia (desire) doubled’ becomes erôs, ‘erôs doubled’ becomes mania. A great many cases in Greek literature illustrate the idea that an excess of desire and love may issue in madness and, more importantly, uncontrolled behavior. Perhaps the paradigmatic case of tragic desire-driven madness is Euripides’ Hippolytus. In the play, Phaedra suffers from an excess of desire: she has been made to fall in love with her son-in-law, Hippolytus. This love issues in irrational and detrimental ‘mad’ behavior: she has ‘wandered from the path of good sense,’ and uses unreasonable, ‘whirling words,’ and eventually even commits suicide.

---

1 Phaedo 81d, 65b.  
2 Ibid, 65c.  
3 Phaedrus 237d.  
4 Republic 573a – 575e.  
5 Prodikos, fragment 7, in Freeman, Ancilla to the Pre-Socratic Philosophers (Cambridge, MA: Harvard University Press, 1948), 140.  
Thus, on an account in which the good life involves the pursuit of wisdom, irrational desire represents a dangerous distraction, and Plato advocates its rational control. He explains: ‘the soul of a man within him has a better part and a worse part, and the expression ‘self-mastery’ means the control of the worse by the naturally better part.’

The Buddha’s ‘Worldly Desires’

The Dhammapada reveals a conception of desire which is commensurable with that of Plato. First, desire is associated with ‘the physical.’ Like Plato, it is described as consequent on the experience of bodily pleasures: in the analysis of the (dependent) origination of suffering, desire arises from sensation, and the Dhammapada asserts: ‘To creatures happen pleasures and wide-ranging endearments. Hugging these pleasures they hanker after them.’

The Buddhist texts repeatedly stress that desire arises from, and is an attachment to, the world: ‘tied up with passion’s delights, [it] culls satisfaction now here now there.’

In addition, and like Plato, these ‘worldly desires’ represent a dangerously potent source of motivation for wrong objects and actions. Described as ‘net-like’ and ‘poisonous,’ they drive us to seek their satisfaction: ‘Those who are slaves to passions follow the stream (of craving) as a spider the web which he has made himself.’

Further, the penalty for leading a desire-driven life transcends even the seven-hundred and twenty-nine times less happy mortal life described by Plato. On the one hand, the devotee of desire is miserable in this life: ‘Whomsoever this fierce craving, full of poison, overcomes in the world, his sorrows increase like the abounding birana grass.’ On the other hand, desire also traps its devotee in the cycle of reincarnation. On some constructions, the restraint of desire is even described as the necessary and sufficient condition of attaining freedom: ‘Craving’s utter ending, utter stopping, is nibbana. Thus become

Medea’s actions in her ‘love-maddened heart (mainomenai kradiai).’ See also the Iliad, as Ares’ great desire to avenge his son leads him to defy Zeus's wishes. As a result, Athene says, ‘Madman (mainomene), mazed of your wits, this is ruin!’ In a similar way, Achilles’ love for Patroklos and excessive desire to avenge his death continues even after the funeral games. While others sleep, he repeatedly drags Hektor's body around Patroklos’ tomb - clearly irrational and inappropriate behavior. His recurrent urge to abuse Hektor angers the gods: ‘in your heart's madness (mainomenêsin) you held Hektor beside the curved ships and did not redeem him...’ Iliad XXIV: 135 – 136, XV: 128.

1Republic, 573 – 575.
3Samyutta-nikaya, xxii.90, and Dhammapada, XXIV, 8.
4Also see, for example, Dhammapada, I, 20, IX, 11, and XX, 12.
5Ibid., XIV, 2, and XXIV, 14. See also: ‘The craving of a thoughtless man grows like a creeper. Like a monkey wishing for fruit in a forest he bounds hither and thither.’ Ibid., XXIV, 1.
6Ibid., XXIV, 3.
cool, that monk, no more reborn, no more becomes. Beaten is Mara. He’s won the fight, escaped all more-becomings.¹

While it is not the case that the Buddhist texts join Plato in explicitly calling desire ‘irrational,’ it is the case that they assert that one who undertakes a life spent in pursuit of the satisfaction of desire does so out of an essential ignorance – a mistaken assessment of the value of this world and its pleasures which the Buddhists hasten to correct. As the Dhammapada has it: ‘They who imagine truth in untruth and see untruth in truth, never arrive at truth but follow vain desires.’²

Thus, the conceptions of the desires on which the Platonic and Buddhist condemnations are grounded are commensurable. On both accounts, desire is a powerful source of motivation toward bodily pleasures: dangerously prone to intensification and typically motivating exactly the wrong actions, it must be controlled in order to pursue the good life.

**Desire in the Platonic Good Life**

While both the Platonic and the Buddhist texts present accounts on which desire represents a real threat, the accounts of the proper role of desire in the individual’s pursuit of the good life differ.

Perhaps because a complete description of the soul would be ‘a long tale to tell,’ in the Phaedrus, Plato illustrates its familiar tripartite nature and function by comparing it to a winged chariot. On this account, the charioteer represents rationality – he is the ‘the soul's pilot,’ who orders the other parts of the soul in pursuit of our ‘proper food’ and ‘natural’ goal of wisdom. An ignoble horse, by contrast, reflects his view of the desirous part of the soul: while it provides motivational energy, it is a ‘a massive jumble of a creature’ - ‘unruly,’ ‘hot-blooded,’ and ‘hard to control.’³

Consideration of the Phaedrus chariot-soul image reveals both an assessment of the role of desire in the good life, and an understanding of desire well in line with contemporary action theory. On the one hand, the account suggests that actions motivated by desire reflect the direction of the chariot by the ignoble horse. However, the chariot-image effectively dramatizes the mistake with such an arrangement: it is clearly the case that it is the nature of the functioning chariot to be controlled by the charioteer, not by its most unruly and ill-bred horse. At the same time, and on the other hand, the account suggests that desire represents the source of motivation for all actions – including the individual’s pursuit of the ‘soul's proper food.’ Describing desire as a ‘stream’ which may be ‘diverted into another channel,’ he indicates that desire

---

¹Udana, or the Verses of Uplift, in _The Minor Anthologies of the Pali Canon_, II, translated by F. L. Woodward, 40, in Radhakrishnan and Moore.

²Dhammapada, I, 11.

³Interestingly, while a second, ‘thumos’ horse is also described, he receives very little treatment in the text. He is truly ‘modest’ - like the auxiliary in the Republic, he is represented – if at all – as the ‘muscle’ of the charioteer, who achieves glory by carrying out the dictates of reason.
may also be thought of as an all-purpose spur to action which may be directed in any number of ways.\(^1\) While the charioteer is clearly intended to be in charge, he would be immobile without the strength of the horses. Martha Nussbaum’s analysis of the import of the *Phaedrus* image is apt: by itself, reason is a ‘relatively impotent moving force,’ as the charioteer requires the energy of the horses.\(^2\)

As such, the *Phaedrus* chariot-image highlights the difficulties with Plato’s account. On the one hand, desire aims at the wrong things, while, on the other hand, reason is powerless with it. Plato seems to be suggesting that desire is both detrimental and necessary for actions in the good life.

Clearly, there is some conceptual confusion begging for resolution here. As Wayne Davis notes, this is not uncommon; the use of ‘desire’ is quite often ambiguous. In his ‘The Two Senses of Desire,’ he makes a useful distinction between the appetitive desires and volitive desire. On his account, the phrase ‘appetitive desires’ is best understood as a noun, referring to an appetite, hungering, craving, yearning, longing, or an urge.\(^3\) On the one hand, in the narrowest sense, it refers to an inner force, the ‘desire to consume things like food and water or to engage in activities like eating, drinking, and sex.’ On the other hand, in a wider sense, many appetitive desires are not appetites, but rather urges toward objects which are ‘appealing,’ or ‘viewed with pleasure.’\(^4\) His phrase ‘volitive desires,’ by contrast, is best understood as a verbal form, referring to wanting and wishing.\(^5\) Davis asserts that, to desire in this, volitive way, is to intend, to decide, or to will. These desires, then, are distinguished from appetitive ones by their ‘rationality’:

Volitive desires are typically based on reasons. Your reason for wanting to eat may be to get nourishment. Then your desire to get nourishment both explains and motivates your desire to eat. Appetitive desires, in contrast, are not the sorts of thing we have reasons for or against. There may be reasons why we have a desire to eat (such as food-deprivation), but we do not have reasons for having a desire to eat. ‘Why do you have an appetite?’ is a legitimate question, ‘What are your reasons for having an appetite?’ is not. Appetitive desires can be explained but are not motivated. In this respect, appetitive desires are more like aches and pains, while volitive desires resemble beliefs.\(^6\)

While the two types of desires are ‘logically independent,’ they are not necessarily also causally independent. Appetitive desires guide actions by motivating volitive desires: my appetitive desire for food inspires my volitive

---

\(^{1}\) *Phaedrus*, 485d.

\(^{2}\) Martha Nussbaum, *The Fragility of Goodness* (Cambridge: Cambridge University Press, 1986), 214. *At Phaedrus*, 247e, the charioteer is even described as nourishing the horses.


\(^{4}\) Ibid., 65 – 66.

\(^{5}\) Ibid., 63.

\(^{6}\) Ibid., 70.
desire to dash off to Little Italy for dinner. However, the link is not necessary: I can have the appetitive desire for food, but, restrained by my recollection of having eaten a large lunch, not form the corresponding volitive desire to head to Little Italy. Thus, as Davis asserts, ‘if a person with a desire to do something does not want to do it, there must be some countervailing factors.' Further, the stronger the appetitive desire, ‘the harder it is to resist, and so the less likely it is that any given countervailing factor would be effective.’ These countervailing factors are often products of the rationality of volitive desires, and, especially, their responsiveness to value judgments:

Value judgments have a powerful influence on volitive desires. A person who believes that something is or would be good, right, obligatory, just, prudent, and so on, tends to want it to exist; he at least has a reason for wanting it to exist. Indeed, volitive desires can almost be identified with one sort of value judgment: The volitive desire to do A tends to result from the belief that it is desirable to do A, which is implied by the belief that one should or ought to do A, that A is the (best) thing to do. 2

Davis’ terminology is useful in resolving the tension in the Platonic account of desire. On this reading, brute appetitive desire (the horse), unchecked by volitive desire, is problematic. Unreasoning, and tending to excess, it is unconcerned with value judgments, frequently motivates an individual to aim at unbeneﬁcial, physical pleasures, and so is often the irrational source of inappropriate actions. Volitive desire (the chariot under the direction of the charioteer), draws on the motivational power of the appetitive desires, but is guided by value judgments, and may rationally motivate us to pursue wisdom and the good life, and so be a stimulus for appropriate actions. 3

Interestingly, when Plato’s account is understood in this way, it is revealed as an anticipation – rather than a rejection – of contemporary action theory. Plato joins Audi in asserting that we should understand an individual’s intentional actions ‘in terms of wanting and believing.’ Even the stronger claim that desire is a necessary condition of action – that ‘intending entails wanting’ – fits neatly with Plato’s characterization of the desire-horse as the

---

1Ibid.
2Ibid., 71.
3This also helps make sense of some puzzling bits of the Republic, where he seems to assert that each faculty of soul has its own desires: the appetitive part of the soul’s desire for money is an appetitive desire, while the rational part of the soul’s desire for wisdom is a volitive desire. (Republic 581a-b) These ‘volitive’ desire for wisdom are even described as capable of serving as a real source of human motivation: ‘…it was the nature of the real lover of knowledge to strive emulously for true being and that he would not linger over the many particulars that are opined to be real, but would hold on his way, and the edge of his passion would not be blunted nor would his desire fail till he came into touch with the nature of each thing in itself by that part of his soul to which it belongs to lay hold on that kind of reality.’ (Republic 490 a-b).
4Audi, ‘Intending,’ 395. Wayne Davis’ construction maps on even more easily: ‘intentional actions are to be explained by reference to the beliefs and desires of the agent.’ (Davis, 10)
source of motivation. Thus, behind Plato’s ostensible condemnation of desire, lies an understanding of desire which is well in line with contemporary action theory, and an assessment of its role in intentional action which (unsurprisingly!) emphasizes rational command over irrational urges.

This reading of Plato is played out in the *Phaedrus*’ seemingly bizarre approval of desire-driven madness. Rather than being extirpated, appetitive desire should be retrained so that it serves as a ‘fund of energy’ which supplies the volitive desires, rather than its own. Even further, the *Phaedrus* suggests that carefully-controlled desire may have an important place in the best life - as useful motivation to a life spent in pursuit of wisdom, it should be allowed to increase in order to attain the best human life.

Drawing on Prodicos’ formulation, Plato even recommends a divine erotic madness – a type of intensified desire which, when for an appropriately philosophical beloved and in an appropriately philosophical love relationship, can provide the impetus needed for the long journey to wisdom. Desire serves to direct the soul’s attention - it can identify beauty in the beloved, the form which is ‘most manifest to sense and most lovely of them all,’ much better than reason can pick out the virtues, which lack ‘luster.’ Further, its strength provides motivational energy for movement toward the beloved, and so toward the forms – it contributes to the doubling of *erōs* which issues in the now-useful mania.

On this account, then, the best life is that of a ‘seeker after wisdom or beauty, a follower of the Muses and a lover...’ - a full life spent in the pursuit of truth, but one motivated by desire-driven divine mania for an appropriate beloved. Thus, the *Phaedrus* issues in a view on which appetitive desire, when checked by volitive desire, issues in the best life:

If the victory be won by the higher elements of mind guiding them into the ordered rule of the philosophical life, their days on earth will be blessed with happiness and concord, for the power of evil in the soul has been subjected, and the power of goodness liberated; they have won self-mastery and inward peace. And when life is over, with burden shed and wings recovered they stand victorious in the first of the three rounds in that truly Olympic struggle; nor can any nobler prize be secured...

Thus, Plato’s account of action is not in conflict with a theory like Audi’s, and his condemnation of desire is far from unequivocal.

---

1Ibid., 391. My underlining.
2F. M. Cornford, in T.M. Robinson, *Plato’s Psychology* (Toronto: University of Toronto Press, 1970), 56. In his discussion of philosophical (or: ‘Platonic’) love, Socrates provides an account of the training of desire, on which the charioteer repeatedly (and harshly!) restrains the ignoble horse until it is controlled: ‘And so it happens time and again, until the evil steed casts off his wantonness; humbled in the end, he obeys the counsel of his driver, and when he sees the fair beloved is like to die in fear...’
3*Phaedrus*, 250d.
4Ibid., 256b.
The Buddhist Good Life

While Plato provides an account on which volitive desire - and even appetitive desire, when under appropriate control – can serve as an impetus for action in the good life, the *Dhammapada*, provides a thoroughgoing condemnation of desire. For the Buddha, desire is a phenomenon which, while motivating intentional actions, arises as a result of ignorance, causes suffering, and so must be completely extirpated.

The Buddhist condemnation of desire arises from its analysis of the human condition. The foundational texts of the Buddhist account clearly associate the good life with the end of the suffering which is held to characterize human life. As they identify desire as the cause of suffering - ‘from desire arises grief, from desire arises fear. To one who is free from desire there is no grief’ – they prescribe the elimination of desire from the good life.¹

On the one hand, and at the most basic level, the texts make the straightforward claim that desire, ultimately, leads to suffering. While desires which go ungratified cause suffering in the obvious way, even desires which we do manage to gratify simply lead to more desires. As the *Dhammapada* explains: ‘There is no satisfaction of one’s passions even by a shower of gold pieces. He who knows that ‘passions are of small enjoyment and productive of pain’ is a wise man.’² On the other hand, and at a deeper level of analysis, the Buddhists hold that the phenomena of desires reflects ignorance of the truth that the self, like the rest of the universe, is transitory. Like Plato, the Buddhist sage Nagasena employs a chariot-image to illustrate the nature of the self:

Just as the word ‘chariot’ is but a mode of expression for axle, wheels, chariot-body, pole, and other constituent members, placed in a certain relation to each other, but when we come to examine the members one by one, we discover that in the absolute sense there is no chariot… In exactly the same way the words ‘living entity’ and ‘ego’ are but a mode of expression for the presence of the five attachment groups, but when we come to examine the elements of being one by one, we discover that in the absolute sense there is not living entity there to form a basis for such figments as ‘I am,’ or ‘I.’³

On this no-self account, then, the pursuit of one’s desires is destined not only for failure and frustration, but also for an essential incoherence. As the *Dhammapada* has it: ‘The fool is tormented thinking ‘these sons belong to me,’ ‘this wealth belongs to me.’ He himself does not belong to himself. How, then, can sons be his? How can wealth be his?’⁴ It should be noted here that, since it is the condition of desiring that is the problem and not just some particular type of desire or particular class of desired objects, it must be desire itself which is problematic. As such, the Buddha advocates the destruction of all desire.

¹*Dhammapada*, XVI, 8.
²Ibid., XIV, 8.
³Visuddhi-magga, in *Buddhism in Translation*, 132 – 5, in Radhakrishan and Moore.
⁴*Dhammapada*, V, 3.
texts are clear that this is the prescription: the true Brahmin is he ‘who has no desires for this world or for the next, who is free from desires and who is separated from impurities.’

While this prescription appears to be straightforward enough, a puzzle lurks here. Richard Reilly constructs it in this way: ‘Buddhism has as one of its chief aims the ‘cessation’ of desires and desiring… However, there is a very real paradox (‘the paradox of desire’) here, since one who aims at liberation from suffering can become desireless only if he should so desire; but, in so far as one desires to eliminate all desires, desire can never be eliminated.’ Two interesting things are revealed by the identification of this paradox. First, it reveals the presupposition that the action of eliminating desire can only be undertaken if it follows from an earlier desire. The texts themselves concur in this understanding: the wise man is described as following the ‘right desires,’ and as ‘always putting forth strenuous effort,’ being ‘evermindful,’ vigilant and earnest. As such, and again like Plato, the Buddhist account is well in line with an action theory like Audi and Davis on which desire is necessary for action. Second, it reveals a particularly vexing prescription: that we should desire to be desireless. John Visvader usefully compares it to ‘an intention paradox’:

‘Raise your hand without intending to’… To raise my hand in attempting to obey the instruction I must intend to do so, yet if I do it with intentionality, I will not be obeying the instruction… I cannot begin or even try to give up all desires unless I first have the desire to do so, and if I take the statement seriously I should give up that desire as well. So even if it was clear what was meant by giving up desires, I must begin by adding to them.

Actually, stated in this way, this is not necessarily a problem – as Wayne Alt points out, it is not necessarily incoherent to begin the process of eliminating desires by adding to them (he compares it to beginning the process of leaping up into the air by bending one’s knees down towards the floor). However, while the project may thus be understood as coherent, it still appears to be impracticable: even if the (volitive) desire to end all appetitive and other volitive desires is met with success, you still have that volitive desire, and so have failed. Bart Gruzalski suggests that this does not necessarily follow: ‘a person could desire to rid herself of all desires except the desire to do so. Once all other desires were extinguished, the desire to be rid of them would be

---

1Ibid., XXVI, 28. See also: ‘a priest who knows things as they really are, he is on the road to an aversion for things, to absence of passion for them, and to cessation from them.’ Visuddhi-magga, xviii,132 – 5.
3Dhammapada, II.
5Wayne Alt, ‘There is no paradox of desire in Buddhism,’ Philosophy East and West 30, no. 4 (October 1980), 521 – 528.
satisfied, and hence it too would be extinguished.'¹ But, as we have already seen, and as Gruzalski himself notes, the idea that the satisfaction of desires leads to their elimination is ‘one of the illusions about reality that Buddha attempted to correct,’ so that this response may provide a logical, but not necessarily a practical, solution.

An alternate approach has it that one can eradicate desire by means which are not necessarily intended to eradicate desire – for example, by the elimination of ignorance from which desire arises. As the Visuddhi-magga has it: ‘a priest who knows things as they really are, he is on the road to an aversion for things, to absence of passion for them, and to cessation from them.’² Gruzalski concurs: ‘It is plausible to think that the person who has an awareness of the nature of all phenomena, including their essential impermanence, will not have any occurrent desires again…Hence there need be no desire to remain desireless…For such a being there will be no desires or attachments of any kind at all.’³ However, as Reilly points out, this understanding may be a necessary, but not a sufficient, condition of desirelessness – I, for one, am able to pursue the fleeting pleasures of chocolate cake, in full knowledge of its, and my own, transitory nature.

Another alternate approach suggests that one can eliminate desire by means of the practice of meditation. As Visvader explains: ‘Special mental spaces are aimed at in which the craving or desiring mind no longer has relevance and ceases, for a time, to function. This alternative mode of using the mind begins to develop its own momentum and gradually expands to cover activities outside the meditation room. The more this kind of space expands in the student’s life, the more the desiring mind is worn away.’⁴ Still, even if the use of meditation solves the practical issues, there are other conceptual issues which remain. If it is the case that we are operating on a view that joins Plato and Audi in understanding desire as necessary for intentional actions, we are left wondering how we should comprehend the actions undertaken by an enlightened individual. Three possibilities present themselves. First, perhaps it is part of the view that either it is – or becomes - possible to undertake actions in the absence of desire. Interestingly, this has a striking corollary in the Platonic account: while gods surely act intentionally – and live the good life of wisdom - Plato describes their souls as lacking desire – they have only ‘good’ or ‘well-matched’ horses.⁵ Second, perhaps it is instead the case that the actions undertaken by an enlightened person are not intentional. This seems to fit with the suggestion made by Buddhist references to ‘skillful actions’ which are not motivated by desire for any particular end.

²Visuddhi-magga, xviii, 132 – 5.
³Gruzalski, 64 – 5, quoted in Reilly.
⁴Visvader, 462.
⁵Phaedrus 247d. In a similar – albeit odd - way, Socrates discusses a myth on which cicadas were released from the demands of the bodily appetites to ‘sing from the very first’ as a ‘boon’ granted by the gods. (Phaedrus 259c)
Finally, it is perhaps the case that the Buddhist account simply does not allow for a coherent view of individual intentional actions without desire. I find this third possibility to be the one that is most likely intended by the Buddhists. If the belief in the self reflects an essential ignorance and leads to sorrow, a prescription which precludes an understanding of the self’s agency can only conduce to the good life. As the Dhammapada itself has it: ‘He whose passions are destroyed, who is indifferent to food, who has perceived the nature of release and unconditioned freedom, his path is difficult to understand like that of birds through the sky.’

Conclusion: Action without Desire?

Plato and the Dhammapada present accounts on which desire represents a real threat – dangerously prone to intensification and typically motivating exactly the wrong actions, it must be controlled in order to pursue the good life. Further, both the Platonic and Buddhist analyses have issued in an account which is surprisingly well in line with that of contemporary action theory. Both anticipate Audi, Davis, and Davidson’s understanding of desire as motivation, and may even happily join them in conceiving of desire and belief as leading to intentional actions.

However, at the same time, the condemnations are by no means identical. While Plato provides an account on which volitive desire - and even appetitive desire, when under appropriate control - can be used as a source of motivational energy in the good life, the Dhammapada calls for the elimination of desire itself. For Plato, desire is an irrational and dangerous part of the soul which may be harnessed for the individual’s pursuit of the good life, while, for the Buddha, desires are phenomena which arise as a result of ignorance, cause suffering, and so must be completely extirpated. While Plato’s seeming condemnation of the place of desire in the good life boils down to a prescription to beware the dangers of unchecked appetitive desire, the Buddha’s call for a total re-imagining of the nature of actions in the good life.

Selected Bibliography


1Dhammapada VII. See also: ‘Now are you seen, O builder of the house, you will not build the house again. All your rafters are broken, your ridgepole is destroyed, your mind, set on attainment of nirvana, has attained the extinction of desires.’ (Dhammapada XI).
2I am grateful to the participants in the ATINER 2007 International Conference on Philosophy for insightful comments on this paper. In addition, I would like to thank the faculty and students of the College of Wooster Philosophy Department for valuable discussions of these ideas, and the College of Wooster Faculty Development Committee for supporting my attendance at the conference.
‘Weeds are the Bane of Fields; Desire is the Bane of Mankind’: Desire, Intentional Actions, and the Good Life


What are the Prospects for Armchair Ethics?

Michael Winter, University of St. Thomas, USA

A priori or armchair philosophy certainly has its place in human inquiry, but it is worth thinking about whether the discipline of ethics can be adequately conducted the armchair. The paradigm armchair ethical theorist was Kant, and his influence on the field is nothing short of astounding. 1 This alone gives us good reason to take the prospects for success of armchair ethics very seriously. More support for the armchair approach comes from considering the nature of thought experiments in the literature of ethics featuring creative, if not sometimes bizarre, situations involving identical twins headed for waterfalls, fat men stuck in caves, villagers lined up for execution, hoodlums setting cats ablaze, trains headed toward exotic sports cars, violinists attached to one’s person intravenously during the night, etc. The not uncommon practice among professional philosophers of contriving thought experiments to demonstrate key ethical points reinforces the perception that ethics is properly done from the armchair. In addition, our reactions to such stories are often based on our ethical intuitions, which some associate with a priori reasoning.

Although I think that a priori reasoning has a place in ethical enquiry, I think many philosophers have underestimated the role empirical considerations can and should play in well-rounded ethical theorizing. 2 After considering two perspectives illustrating armchair approaches to ethics, I point out some limitations of this way of doing ethics. I begin by focusing on a case made by Richard Fumerton, who argues for the thesis that the distinctively philosophical work in ethics is best done from the armchair. I then focus on Michael Smith’s work because I think it presents a concrete example of an attempt to build an

---

1 Consider Kant’s influence in the work of Moore, Prichard, Donagan, Gewirth, Nagel, and Korsgaard, just to name a few.
2 Of course, many do emphasize the importance of an empirical approach. Among them are Richard Boyd, Gilbert Harman, Nicholas Sturgeon, David Brink, and John Doris.
ethical theory from *a priori* foundations. Some of the criticisms I offer draw attention to certain features of Aristotelian virtue ethics that I think are not subject to the same shortcomings. Although it would be another project to develop the case for a virtue theory understood in this way, I will make some brief suggestions throughout this discussion about how this might be done.

**Fumerton’s Case for Armchair Ethics**

Richard Fumerton argues that distinctively philosophical ethics is done from the comfort of the armchair. He first distinguishes between normative ethics and metaethics, and claims the latter deals with ‘questions that concern the analysis of fundamental ethical concepts such as good, bad, right, wrong, should, shouldn’t, and a host of concepts… probably definable in part by reference to our paradigm ethical concepts.’¹ Having provided this characterization of what a metaethical concept is, he argues that such concepts are properly analyzed from the armchair. Fumerton admits that some normative questions in ethics are probably not answerable from the armchair, but suggests that departure from the armchair to answer such questions is tantamount to taking off one’s philosophical hat.

Fumerton’s argument for the conclusion that genuine philosophical ethics is best done from the armchair rests on the premise that genuine philosophical ethics, i.e. metaethics, consists exclusively of *a priori* content. Part of the case for this premise is based on how ‘metaethics’ is defined. The definition Fumerton offers (quoted above) or some slight variation of it is used widely, so it should not raise many eyebrows. But this definition is closely related to the substantive claim that ‘questions that concern the analysis of fundamental ethical concepts such as good, bad, right, wrong’ are all *a priori* in nature. The additional bite in the premise comes from the idea that normative ethics is not genuinely philosophical. Let us consider the case offered for each of these ideas.

The support for the idea that metaethical concepts are *a priori* comes from considering the extent to which metaethical questions are dealt with by means of conceptual analysis. Consider the is/ought problem, a popular and often discussed metaethical puzzle. If the is/ought problem is thought of as raising questions about how it is possible to derive evaluative conclusions from factual, or non-evaluative premises, then surely attempts to resolve the is/ought problem would involve considering the nature of inference, possibly whether there is any difference between logical and practical inference, what is involved in making an evaluative claim, and so on. All of these concepts can be clarified by abstract thought alone, and it is not clear how empirical considerations would have any bearing on them. Or consider another central question in metaethics—how we reconcile the apparent objectivity of ethical judgments with their acknowledged practical force? Dealing with this question involves a lot of hard thinking about the standards used to determine whether certain kinds of claims are true, the nature of motivation, desire, etc. Again, it

is not clear how empirical considerations would be very helpful in working through issues like these. It does seem as though these are armchair issues if there ever were any.

What supports Fumerton’s thought that normative ethics is not genuinely philosophical? Here Fumerton embraces the positivist idea that normative questions are solved by sorting through the metaethical issues, plugging in the empirical data, and then seeing what follows. Philosophical inquiry is surely relevant at the metaethical level, but he notes how ‘plugging in the data’ is probably best done by people with good common sense, a commodity that may be hard to find among any given group of philosophers, but is more easily found among practitioners of various fields where specific normative questions are relevant.

What should be said about this argument, particularly about the two claims supporting the premise that genuine philosophical ethics, i.e. metaethics, consists exclusively of \textit{a priori} content? I think both supporting claims are dubious. First, I am not convinced empirical considerations have no bearing on metaethical questions. I am willing to grant that the is/ought relation and the problem about the objectivity and practical force of moral judgments are armchair issues, but there are other metaethical concepts upon which empirical considerations seem to bear. Consider, for example, Bernard Williams’ criticism that consequentialism forces us to abandon our personal projects, thereby creating a sense of alienation. We might think this point has some degree of plausibility, but because of its abstract nature be unsure about whether it is true. In order to find out whether consequentialism leads to the type of alienation Williams describes, we might have to think as a consequentialist would in moral situations we face in our own lives. Because this process of determining the truth of the meta-ethical principle would be tied to specific cases we may have faced, it is not clear how this type of thinking would be \textit{purely a priori}. Even if armchair or \textit{a priori} considerations about the truth of the meta-ethical principle are what is decisive in our determining whether to accept the principle, it nonetheless seems to be the case that these armchair reflections will only be plausible to the extent that they conform with what we have experienced in concrete moral situations in our lives. Does our moral experience support the idea that the type of alienation Williams describes comes with consequentialism? If it does, then this empirical evidence supports the argument. If it does not, then empirical evidence counts against the argument. In this instance we can see empirical considerations grounded in our own moral experiences can be quite relevant in helping us determine whether a metaethical point is legitimate. If this is correct, then metaethics may not be the purely armchair endeavor that Fumerton makes it out to be. There will be many other instances where empirical considerations will not be relevant to a metaethical principle, e.g. can an evaluative claim be entailed by a factual claim? But all we need is one plausible case that illustrates how empirical considerations would be relevant to show that metaethics is not thoroughly \textit{a priori}.

\footnote{Fumerton (1999) p.31.}
Concerning the question about whether philosophers have anything distinctive to say about normative questions, I think we can see that they can when we consider concrete issues in normative ethics. Consider the abortion issue, for example. Surely philosophical considerations are useful in helping us see whether human life has some special kind of dignity. We might think determining whether and when this special property exhibits itself in the continuum between birth and death is a matter answered by appeal to empirical considerations. If so, we would expect that doctors or possibly embryologists should have something decisive to say about such matters. But we are likely to discover that the level of controversy and disagreement among scientists and medical practitioners about the moral status of the unborn is as about as high as it is in the general public. Why? Because empirical considerations do not by themselves settle the issue, and a sometimes complicated interplay between empirical and conceptual issues is what often separates people on such questions. Philosophers can help sort through the conceptual quagmire, even if they cannot provide a final answer on the question that will be convincing to everyone. Philosophers do have something valuable to contribute to these discussions. So I do think philosophers have quite a bit they can offer on the normative front. I will say more about this point below, but for now it seems fair to think that if philosophers have something distinctive to offer on normative issues, and if what I say about how empirical considerations might help us understand some metaethical issues is correct, then Fumerton’s case for the claim that genuine ethics consists of exclusively a priori content is not very strong. This should lead us to be cautious about what we think can be done in ethics from the armchair.

Michael Smith’s Dispositional Theory of Value

Let us now shift attention from this general argument about the nature of ethical inquiry to a specific proposal about how we might go about framing a moral theory a priori. Michael Smith attempts to construct a model for ethical theory that incorporates realism, naturalism, and internalism. One attractive feature of Smith’s approach to moral theory is that is has something to offer nearly everyone, although most of its offerings will be treasured by moral realists. For example, Smith takes very seriously ideas often associated with Kantianism—establishing an a priori rational foundation for ethics that is categorical and not merely hypothetical. Most interestingly, Smith’s theory is at once both Humean and anti-Humean (in different respects of course!). Smith is inspired by the Humean belief/desire paradigm of moral psychology, but attempts to construct an anti-Humean, non-relativist realist ethics that is built on the belief/desire paradigm.¹ What is more, Smith’s theory is a dispositional theory of value, which gives it a bit of Aristotelian flavor—it frames value in terms of what we would be disposed to value under ideal

¹Dancy (2004) calls the position ‘Humean Realism.’
What are the Prospects for Armchair Ethics?

Smith says ‘facts about what there is a justification for doing, in various circumstances, are in turn plausibly thought to be facts about what we would advise ourselves to do if we were better placed to give ourselves advice…’2 Those familiar with the role of the phronimos (or practically wise person) in Aristotle’s ethical theory, will probably notice a striking affinity on this point.

Smith’s theory is based on the principle that ‘what is desirable for us to do in circumstances C is what we would desire that we do in C if we were fully rational.’3 This principle offers a potential basis for moral realism because according to it, we have a way of saying that propositions with moral content in fact true.4 The standard for determining whether or not such propositions are true is the justification one would offer under idealized conditions of reflection5, where the idea is that an agent is able to determine the morally appropriate thing to do in any set of circumstances by thinking about what she would advise herself to do if her desires were immune to rational criticism, i.e. if her desires formed a set that is maximally informed, coherent and unified.6 If an agent sees this justification, then she will have the desire to perform the prescribed action all things being equal—absent depression, weakness, or some other form of irrationality.

This theory promises to account for the objectivity of moral judgments because it grounds the justification for what is right to do in terms of rationality as such. In this way Smith’s theory differs substantially from other internalist models like those offered by Bernard Williams and Gilbert Harman, both of whom offer variations of a relativistic internalism. Since rationality as such is still a subjective feature in some sense—a feature of subjects—moral injunctions will have prescriptive force as well. By accounting for both the objectivity of moral judgments and their prescriptive force, Smith claims to have solved what he believes to be the most significant problem in ethical theory.

What Smith says about externalism gives us insight into his views about how ethics should be done. He notes that one problem externalists face is squaring an empirical methodology with the apparent fact that ethics seems to be an a priori endeavor. If so, how can we make sense of the idea of interacting causally with a priori principles? Smith is straightforward in characterizing ethics as an a priori endeavor. He says: ‘…it is agreed on nearly all sides that moral knowledge is a relatively a priori matter, at least in the following sense: if you equip people with a full description of the circumstances in which someone acts, then they can figure out whether the person acted rightly or wrongly by just thinking about the case at hand. Someone who claimed that it would be impossible to figure out what is right by just thinking about

---

1When I claim this is an Aristotelian idea I have in mind the fact that Aristotle seems to be the first moral theorist to have placed importance on the idea of dispositions of an agent under ideal circumstances.
circumstances of action would be mis-using the word “right.”¹ One thing to be noted is that the remarks here seem to refer to normative ethics, because the focus is on whether someone acts rightly in a described set of circumstances. Fumerton would classify this as a normative issue at any rate. Smith’s remarks seem to apply especially well to the types of thought experiments mentioned earlier.

Consider Peter Singer’s proposed solution to world hunger as an illustration of Smith’s point.² The alleged solution rests on an argument that makes a strong appeal to our intuitions about what would be the morally right thing to do in a hypothetical case. Singer uses a colorful story from Peter Unger in which we are in a position to flip a switch so that a speeding train will be taken off course from smashing into our exotic automobile. But a consequence of our action is that the train will run over a helpless child instead. Our intuitions tell us it would be crass to place more value material objects than on human life. From this story Singer hopes to have us take seriously specific conclusions about giving away our excesses to relieve famine in third world countries.

I think this story about flipping the switch to possibly alter the course of a speeding train is a good example of armchair ethical theorizing, and I think it illustrates the sort of activity fitting well with the description Smith gives about the a priori nature of ethics. As I indicated above, I do not doubt that a priori considerations have an important role to play at both the normative and meta-ethical level, but I doubt that this characterization accurately depicts the nature of ethical inquiry generally.

**Worries About Armchair Ethics**

My basic criticism of armchair ethics is that it ignores important procedures for getting information that can help us make progress in understanding ethical concepts. I believe some type of restricted armchair philosophical ethics might be possible, but that we unnecessarily limit what ethics can do when we think about ethics as an exclusively armchair endeavor. A case was made above for the point that empirical considerations seem to bear on at least some important meta-ethical questions. In addition, induction can be seen to yield moral principles often utilized to think about cases in normative ethics. There are facts about ourselves learned inductively, often through lots of firsthand experience, that should not be understood as part of what Smith calls the ‘circumstances in which someone acts.’ This inductive grasp of general moral principles is not to be confused with casuistry, or applying general principles to particular cases. Characterizing philosophical ethics as a thoroughly a priori or armchair discipline fails to leave room for the idea that some general moral principles can or must be acquired inductively, both at the normative and meta-ethical level.

In addition, there are good reasons for thinking that some of our own attitudes are not discoverable a priori, and these attitudes might have an

---

What are the Prospects for Armchair Ethics?

important role to play in normative ethics. Consider the Implicit Associations Test (IAT), which provides a good example of how some of our attitudes do not seem to be discoverable a priori. This test is meant to show, among other things, that we are not always capable of knowing our own minds—that people may be unaware of some of their own attitudes. If the test is capable of showing this, then we have further reason for being suspicious of armchair ethics.

Those taking the IAT can choose from a number of topics to test subconscious attitudes, including: race, age, sexuality, religion, gender-science, and many others. The test begins by placing simple categories on the left and the right of the screen, GOOD on the left and Bad on the right, in the racial test, for example. The user then taps a key on the left of the keyboard when a word generally associated with goodness—words like ‘joy’, ‘wonderful’, ‘happy’, etc., flashes on the screen. A user taps a key on the right when a word appears that is associated with badness—words such as ‘agony’, ‘horrible’, ‘nasty’, etc. The user is instructed to respond as quickly as possible. The next phase of the test uses the categories White American and Black American, and faces or portions of faces of whites and blacks are flashed on the screen. Again, the user taps keys on keyboard to place the images in the appropriate category. The final phase of the test combines the categories, so on the left we might have Good or Black American; on the right we would have Bad or White American. As words and images from any category rapidly appear on the screen, the user has to quickly place the words in the correct set of categories. The speed of the responses and the number of errors together indicate something significant about one’s implicit attitudes about race in the case of this test.

Without being aware of such a test, we might think we could determine introspectively (a priori) what our attitudes about racial groups are by considering our beliefs and carefully attending to how we act and feel in specific situations when we interact with members of other racial groups. Suppose we determine after such introspection that we are basically racially color blind and that we have no negative associations with members of other races. Some of the moral principles we act on when we interact with members of other racial groups seem to be dictated by what we think about our own attitudes. Suppose that (a) I think it is morally good to be racially color blind, (b) I want to be racially color blind, and (c) that introspection tells me that I am. When I interact with members of other racial groups I will simply rely on the same moral principles I use when I interact with members of my own racial group. But now suppose that (a)-(c) hold, but I take the Implicit Associations Test and learn that I have a strong preference toward members of my own racial group and lots of negative associations with members of other racial groups. In view of this discovery, if I strive to be morally good, I may need to invoke additional moral principles when I interact with members of other racial groups, maybe a principle stating that I should make sincere effort to make

---

1The test is on the Web at https://implicit.harvard.edu/implicit/
2Gladwell (2005) describes and explains the test and its implications.
3Gladwell (2005)
kind gestures toward members of other racial groups, for example. Once I learn
that the more exposure I have to the positive achievements of members of other
racial groups tends to lower my negative associations with a group\(^1\), it is
reasonable to think that a person trying to be morally good will strive to make
efforts to find opportunities to do so.

This case shows that the sorts of principles a person who wishes to be good
brings to bear on a situation depends on judgments about situations formed by
experience with facts about the case at hand and facts about ourselves. Such
facts are not reliably accessible in an *a priori* manner. And this is a significant
aspect of normative ethics.

Furthermore, Smith’s standard for determining the right action depends on
the notion of a person situated in ‘idealized conditions of reflection.’ When we
ponder this phrase, it is hard not to conjure up images of a deeply reflective
agent sitting comfortably in an armchair thinking through the details of some
hypothetical situation. But I think the coherence of the concept of such an ideal
requires an ability to discern which principles are most salient to the case at
hand, and this ability would seem to require the sort of habituation and good
character traits that come with lots of experience as a moral agent. Put in
another way, thinking about a situation as a fully rational agent would requires
the sort of perspective on a situation that a fully rational agent has. But this
perspective is not something one gains in the armchair; rather, it is gained by
being engaged in a moral life, which involves making mistakes and learning
from them among many other things. This is part of the inductive grasp of
moral principles mentioned above, something sure to be featured in a well-
developed virtue theory of ethics.

Maybe we should say that it is a mixed bag as far as ethics is concerned—
some of its propositions are empirically based while others are not. But absent
some cogent argument for the conclusion that empirical propositions are not
relevant to the subject-matter, it is desirable to try to make use of empirical
data wherever relevant. There is reason to believe empirical work has much to
lend to ethical inquiry. John Doris, for example, makes a case for the idea that
there are no steady, reliable character traits on the basis of empirical data. In
other words, his work attempts to establish a substantive metaethical
conclusion on the basis of empirical data. Whether Doris’ conclusion follows
or not, it is reasonable to think that some metaethical conclusion or other will
follow from the data he considers together with supplementary conceptual
points. The project is fascinating and its results should be considered carefully.
Doris is right to insist that philosophers look beyond what seems right to them
and look carefully at how their conclusions match up to systematic
observations of behavior.\(^2\)

---

\(^1\) Gladwell (2005)
Virtue Theory as an Option

I have argued that ethics is not an entirely *a priori* enterprise even though *a priori* considerations certainly play a significant role in ethics, particularly in metaethics, but in normative ethics as well. This result could fit quite well within Aristotelian virtue theory developed in a certain way. Aristotle’s own inquiry in the *Nicomachean Ethics* weaves together conceptual analysis with empirical considerations in ways that resonate with many of the points I have addressed. Aristotle’s rich notion of practical reason involves a dispositional analysis akin to the one Smith describes, but the dispositions involved therein are built on enormous experience and attention to specific situations. Furthermore, Aristotle’s theory of practical reason seems to require some type of inductive grasp of ends of action.\(^1\) Of course, realizing the severe limitations of armchair ethics does not force one toward an Aristotelian virtue theory, but given the remarkable level of potential explanatory power such a theory offers, it does seem like a fairly good place to begin.

Bibliography


\(^1\)Dahl (1984) defends this position at length.
Part VI

Philosophy of Region
The Franciscan Concepts of Joy, Laughter and Foolishness in the Post-modern Context

Aleksandr Golozeubov, Kharkiv National University, Ukraine

The laughter of a monk – an excellent subject for thought or meditation, especially in a world tends to snigger because it no longer knows how to laugh. On the usefulness of laughter in human relations. On laughter as a means of relaxation or communication. On laughter as a weapon; everybody knows that it can be implacable. The laughter of the master and the smile of the wise man – a laugh of sweetness, a smile of strength. But also a smile of understanding and reassurance, and of serenity too, the reflection of a very particular sense of whimsical humour, as polished as a pebble worn smooth by the years or centuries, through lifetimes of surge and shift… A grave, fine laughter, for it reflects an echo of plenitude… A smile to heal all the ills of the world, and a laugh to give courage for every day’ (Levenson, p. 27). Unfortunately, a western man mainly lost ability to feel completeness and variety of being and to affirm this feeling in laughter and smile. We address to Buddhism and interpreter it in specific way. Similarly new medievalism represents new vision of the basic values of the western culture. Grotesque and humour have been revealed in the medieval culture and saints announced as holy hippies¹. In looking for the true joy researchers applied to image of St. Francis of Assisi and spiritual tradition, which he founded.

St. Francis was not a unique figure among Catholic saints who talked about joy. But in the post-modern culture his image of God’s Fool has been mainly used for apology of the Christian joy. This joy reflected many basic components of the western culture, actual for the post-modern mind. We follow Umberto Eco in his interpretation of post-modernism, firstly, as certain

spiritual mode, special play, and, secondly, as a phenomenon that has revealed itself in the radical change of paradigms. Such change occurred in 1960th, and not only on philosophical, social, and historical levels, but also on religious and ecclesiastical one (reforms of II Vatican Council). The paradigms have been not only changed but also reconsidered. Christian concepts of sainthood, joy, and laughter have undergone significant transformation and were actualised under influence of liberal ideas and unprecedented interest to play, carnival, and leisure. The joyful spirit has been seen in Christianity, and it stimulated vision of the world as divine comedy, and Bible as the book, which has embodied this paradigm of laughter and humour. Interest to carnival personages and especially to holy fools became the following step. Interest to Franciscans stimulated joyful vision of Christianity, and interest to St. Francis of Assisi to holy foolishness not only in the Eastern Christianity but in the Western one as well. This way of thinking dealt with laughter and buffoonery, and with suffering and martyrdom at the same time. It reflects paradoxical nature of the western thought, which generates both the dance of death and black humour, and the tradition of spiritual joy. In wide sense Franciscan tradition includes a lot of thinkers, poets, artists, martyrs, and missionaries who have reflected first of all Christocentric outlook, for which both joy and passion have had essential meaning.

The issue of joy and laughter for the Catholic in general and especially for medieval consciousness is an unavoidable issue about pain, suffering and death at the same time. In the medieval consciousness violence, non-justice and death was every-day routine, while suffering for the sinners and joy for the righteous men were referred to the apocalyptic future. For St. Francis joy expects us here and now, though through suffering. Francis suffered without violence to himself but he reached the identification with Christ at the moment of His deepest suffering and humiliation that was stigmata. This is actually the theme of the death without the death itself.

Through the story of Christ His followers feel intimate personal attitude to the word, which is in the centre of the created and the imperfect reality. Hence it is clear why St. Bonaventure shares interest in the human nature of Christ overall for Franciscans. He confessed: ‘The focal point of the faith and holiness of St. Francis was his love and imitation of Christ’ (Franciscan theology, p. 2). Christocentric vision of the Franciscan school and doctrine of Christ’s absolute primacy were formulated and expounded by Alexander of Hales and John Duns Scott. Richard of St. Victor, St. Antony of Padua, St. Bonaventure and others who considered the Sacred Heart as the centre of Christ’s Humanity, influenced Franciscan theologian James of Milan. In the preface to his treatise Stimulus Amoris [The goad of love] he wrote: ‘One should willingly give thought to the suffering of Christ. Such meditation on His suffering is very fruitful’ (Franciscan theology, p. 99). Treatise about love and about heart as focus of this love contains the true apology of suffering follows: ‘O lovable suffering...Wanting to sorrow with Christ, one made...’ (Franciscan theology, pp. 101, 103). Further we see absolute confluence with Christ, almost sensual ecstasy: ‘...His wounds are always open, and through these wounds I will enter His womb again, and I will re-open it time after time until I shall be inseparably enclosed within it’ (Franciscan theology, p. 103).
Nevertheless, the full imitation of neither sufferings of Christ nor His foolishness is possible. The reason is not only in the greatest physical and moral tortures. It is important that in the same apology these ones cease to be tortures, demonstrating the ability to exceed the limits of human nature, after coming into contact with a divine sphere and find joy and sweetness in it. But Christ's sufferings were deprived of affectation or sweetness; they were no softened or transformed in any way. These sufferings were renewed with highly dramatic and tragic feelings in the Catholic Church. The apology of suffering in the writing of James of Milan is not about suffering of the martyr. Here actually joy and pain originate from the affected religious feeling being impressed by the same image that the suffering Christ. All this helps to comprehend the God, first of all in His corporal nature. This is not expiatory suffering, but necessary one. In the case of martyr and even more in the case of fool for Christ, this affectation was strengthened by a real danger to be beaten, mocked or killed. The physical suffering, long dying can be experienced no less emotionally and dramatically. Not tragically but just dramatically even before the face of death. The martyr can foresee own tortures and death, but he doesn’t stop. A martyr can even organize his pain in some kind of performance, spectacle. In the thirteenth-century vita of St. Vincent the saint laughs, as he is tortured, his laughter signalling his disdain for earthly trials and his faith in the joys of the kingdom to come. Usually a martyr keeps the image of Christ himself in mind. Martyr St. Ignatius of Antioch in his Epistle to Romans showed his intense love for Christ. He tends to die and to be with Jesus: ‘Let there come upon me fire and cross, and struggles with wild beasts: my being mauled and torn asunder, my bones racked, my limbs mangled, my whole body crushed, my being cruelly tortured by the devil, provided only that I may reach Jesus Christ’ (Treasury of the Catholic Church, 1999, 5:3). Partly a martyr and a holy fool imitated Christ; partly Christ himself inherited similar models of behaviour.

Suffering isn’t good in itself. But it is a necessary and an unavoidable way to good and knowledge. A Franciscan theologian Peter John Olivi reminds in his ‘Letter to the Sons of Charles II’ words from Luke, 24:26: ‘Thus, the meaning of Christ’s world is clear – ‘It was necessary for Christ to suffer and to enter into the glory of his kingdom’ (Apocalyptic spirituality, 1979, p. 176). But it was necessary for Christ to be mocked as well. Similarly it is essential for Christian according to Paul to avoid of the human wisdom of the world, but to be fools for Christ. It means to be foolish in order to be wise. It is far from opposite kind of foolishness that is real stupidity and spiritual deafness.

The distinctive feature of St. Francis’s mysticism is self-identification with Christ, especially in His sufferings, His holy foolishness and childishness, His pleasure and despair. Saint Francis felt special joy during the Christmas but in other cases his joyfulness was inseparable from sorrow. In the image of St. Francis imitation to Christ seems probably to be the most complete and natural. It concentrated both the tragedy and sublimity of crucifix through stigmatisation and mockery, which Christ endured. But Christ didn’t preached to the birds and he didn’t sing. St. Francis had prepared change in the traditional ideas about Christian saint and sainthood, which through several
centuries lead to the radical innovations in the image of Jesus Christ. In the first version of Francis’ life Celano wrote that the saint in his first 25 years was experienced in all kinds of stupidity. He attempted to exceed others in the shameful behaviour; the jokes, the eccentricities, the idle talking, the songs, ever making himself foolish, and the people avoided him. But Francis’ conversion transformed stupidity of his youth into holy nativity and clown into God’s Fool.

Francis can be imagined as the participant of carnival, singing, dancing, and in others this behaviour can evoke holy joy as we can see in Kazantzakis’ novel Saint Francis (1962). This is almost jurodstvo, but the true jurodstvo isn’t carnival, but a provocation. The brothers were singing and were glad together with Francis. His sermons spread with great passion and fervour; and their effect upon people was impressive. Celano who had probably heard Francis preaching, spoke of his introducing ‘ardent gestures and movements’ and on a famous occasion when he preached before the Pope he began to dance, and Cardinal Ugolino, who had arranged for the sermon to be given, was terrified that the Sacred Congregation would laugh. But people were not laughing when Francis was preaching; they were far more likely to cry, as they did in Assisi on the day when he preached naked on the nudity of Christ and the beauty of poverty.

But in the beginning of his religious life St. Francis experienced the mockery and humiliation; therefore perhaps in Christ's perception in the Franciscan tradition, it is this moment that is especially important. In the passion narrative from pseudo – Bonavertu’s meditations on the life of Christ (late 13th cent.) some times it has been outlined that Christ was mocking and was led back and forth, scorned and reproved, turned and shaken ‘here and there like a fool and an imbecile’ (Medieval popular religion, 1997, p. 78). This is rare in Christian literature likening of Christ himself, but not saint who imitates him to a fool. Herod also thinks about Christ as a fool: ‘Thus you see how everyone regarded him not only as a criminal but also as a fool; but he bore everything most patiently’ (Ibid.). Any martyr was perceived rather as a criminal, but not as a buffoon. Buffoon represents feigned death, play, for a criminal it is a reality. But it makes death of the accused martyr the final event. In the case of Christ, death is only the start of the world history. They mocked him as if he wished to rule but did not have the power. Everyone mocked him, ‘not only did he endure their gathering together the whole company for great mockery, but also their leading him before Pilate and all the people to mock him publicly, wearing the crown of thorns and the purple robe’ (Medieval popular religion, 1997, p. 78). Nudity of Christ strengthens his helplessness and both these features are emphasized especially. But for the fool nudity doesn’t stress his ostracism, but ambiguous freedom from the social conventions. For Christ there is no necessity to prove, to show or to imitate something or somebody by his nakedness. This intensifies the feeling of tragedy even more and stresses dramatic nature of the action, description of those tortures, which this body endured, whereas it adds not tragic, but rather comic or farcical element for the fools. They don’t know about the divine nature of Christ and they mock him first of all as a man or a fool. But he is not fool. The fool can’t be God, and God can’t be defenceless and exposed before
the mortal person, hence, Bonaventure calls to look at Christ precisely as a man, moreover, this religious experience which combines joy and suffering is even deeper.

Traditionally Catholicism focuses on Christ, his birth, death and passions and identifies with Christ not in majesty but in passions and sufferings. According to Gerald O’Collins, Catholics want to see not simply a bare cross but a crucifix; they ‘want to see his body and its wounds’ (O’Collins, 2003, p. 369). Nobody approached to the Franciscan imitation of Christ in His passions. But paradoxically this passion doesn’t destroy his joy. Joy should be a transition mode from the human to the divine in order to transform even suffering into the enjoyment. The affectation of death and suffering is replaced by the affectation of complete confluence with Christ, and then by the affectation of joy and happiness. A question is not about why man doesn’t want to share Christ's suffering, but why he doesn’t want to share joy, all the more these sufferings are shared with Christ, and it gives extraordinarily emotional effect. St. Bernardine of Siena described religious feelings of St. Francis in the following words: ‘When he pronounced the word ‘Jesus’ or heard someone say it, he was filled with joy and he seemed to be completely transformed, as if he had suddenly tasted something marvellous or caught the strain of a beautiful harmony’ (English Omnibus, p.710).

The main ideas of St. Francis were systematized, integrated and commented in the Franciscan tradition but also actualised in the modern Church. The concept of joy became one of the key concepts in this context. Though St. Francis had searched for the patronage of the Pope and legalization of the order, the writings of such Franciscan theologians as Peter John Olivi, William of Ockham and others were later subjected to special commissions or even condemned for heresy; first of all in the problem of poverty. But actually the theme of joy, laughter and foolishness in the sermon and life of St. Francis seems to be more new and fruitful.

None of Francis’s followers approached a type of sainthood embodied in St. Francis. He combined an optimistic, joyful perception of life, the adoption of life in its whole complexity, including suffering and pain, a practical character of the doctrine and extraordinary spirituality and mysticism. Besides, he approached to the model of Christ as fool and martyr most nearly though he didn’t inherit it totally. He always longed to die as a martyr. However, St. Francis wasn’t a martyr, he even didn’t imitate Christ, but he followed him, the mode of his life and death up to the stigmata and it did not contain any provocation for him. The Byzantine holy fools seem much more hypocritical and ambiguous, because all way of their lives was provocation. Francis as nobody else imitated life, suffering and the every idea of Christ, adding to his image essential and specific part of his own personality that made possible to call him as God’s Fool. In his Testament St. Francis declared: ‘The ideal friar was to be idiota et subditus omnibus. Once he declared that when a learned man entered the Order he ought to divest himself even of his learning so that he might offer himself naked to Christ’ (Moorman, 1968, p. 54). The followers of St. Francis called themselves the ioculatores dei and mundi moriones. Humour
became a mark of intimacy, and in the willingness to play with sacred institutions we can see easy treatment of religion.

So, St. Francis resembles to us some main components of holy fools’ model, or, in other words, fools for Christ’s sake. I mean self-humiliation – imitation of humiliation of Christ that is a guarantee against pride, the worst sin; nudity that symbolizes the rejection of all social roles; wondering, pilgrimage that means rejection of worldly ties and security. This humiliation in poverty and patience is manifested more strikingly in the chapter from ‘Little Flowers of St. Francis’ about the true and perfect joy.

The Byzantine jurodivyj provokes to be beaten and mocked. He puts the mask of stupidity on himself in order to save other’s soul. He is ‘fool for Christ’, but St. Francis isn’t fool for Christ, he is God’s fool but for the people, there is no provocation in him, but there is a form of frank religious experience. Nobody raises the question about stupidity or real foolishness of St. Francis. He is fool precisely in the religious sense of infantilism, openness, and sincerity.

Holy fool Symeon the Stylite was engaged in the painful battle with his body on the top of the column, the bones and sinews visible on his feet due to the pain. Symeon loved his pain so much that he tortured himself to death (Glucklich, p. 24). Francis didn’t love pain, but he accepted it. He loved people and world, but he understood all mighty of sacrificial love, with which Jesus Christ suffered and died. He assumed the inevitability of suffering in the world and endured pain because it tied him with the world, from one side, and Christ, from the other.

Francis, Dominic and others are the saints, but not martyrs. Though their enthusiasm, persuasion and internal force were not inferior, but rather exceeded the same in other martyrs. They didn’t become outcasts, but they established significant and not just religious, but social institutions within the Catholic Church. Their sermon and method of persuasion and behaviour could seem no less theatrical, than in the Byzantine saints. But it isn’t marked by provocation and ambiguous profanation so much. In the eastern Christian tradition the theme of jurodstvo, buffoonery and carnival behaviour of a saint is more stuck out, though it remains more behind the text. In the Catholicism it exists rather in the latent form, but having been raised, it is experienced with much more dramatic or even tragic effect. Nevertheless, if not obvious buffoonery, then the irony and self-irony are quite inherent in the Catholic saints. The hagiography that depicts a saint as a person telling tasteless jokes, for example, sounds more convincingly than hagiography, which depicts him as only serious one. In the Byzantine saints, the imitation is provocation and a way of stressing the human insignificance. In St. Francis, the imitation is a way of identification and confluence. It potentially contains the senses and possibilities which fools showed publicly, and which became much more provocative in the sense of further development of Franciscan tradition. At the same time Byzantine jurodstvo remained mainly cultural and historical phenomenon.

The issue of laughter and joy in Christianity paradoxically refers to the monasticism and its reception in the medieval society. It seems inevitable, though this is a social institute, which at first sight has little in common with any reflections on laughter. The monk is ‘the person who… wept for his own
sins and for the sins of humanity and who sought, with a life of prayer, meditation and penance, to obtain personal salvation and the salvation of all mankind’ (Le Goff, p. 2). The ‘modesta hilaritas’ was considered as an appropriate mode for the Christians. It implied not simply abstention from the laughter that could be understood as vain and proud attempt to become similar to Christ but absolute non-involvement in it, otherwise a man could be returned into authority of the devil. Medieval theologians thus assumed. At the end a good Christian would hope on the meeting with heaven joy, while sinner committed himself to devil. It meant, according to some theologians, mockery at God, laughing at him and making thus their rescuing impossible. Nevertheless the buffoon, jester and clown are main personages of laughter in the medieval society. They are in the crowd and among the crowd, and they don’t resist it. The martyr is a distinguished and individualistic person. He is above the crowd, on the cross or on the bonfire, but we can hardly consider this place as something like a stage. The fool, the buffoon opens the truth, which is intuitively known to him better than to the crowd. The martyr doesn’t open the truth; he isn’t the conductor of God’s word. He opens himself to God as a victim as if he tells God: ‘Look, how I am dying’. In this sense Christ's death is exceptional, but its uniqueness has been just profaned by the carnivalization of Christ and His identification with the carnival personages of buffoons and fools.

The post-modern novel by U. Eco The Name of the Rose became a litmus paper of this process. This novel belongs to mass culture and at the same time it is an object of serious research studies. Here the theme of medieval laughter and treatise about the laughter is combined with criminal intrigue. This scheme is complicated by the death of the book, when this text of Aristotle about the laughter was burned in the library. Both the publishing of Eco’s novel and its problematics reflect the growing interest in the Middle Ages and its culture, religion and daily life. But these Middle Ages are read in the context of the postmodernism with accent on semiotics, sign itself, meta-narrative, variety of interpretation and obvious play with reader.

Franciscan friars, similar to Dominicans, have remained mainly inside this paradigm but a new kind of mysticism and perception of the nature, born by St. Francis, inspired them on more bright vision of the world around. This spirit could not be saved a very long time, especially after the establishment of Franciscan Order as not homogeneous but an influent religious and social Catholic institution, from one side, and Black Death wave and various social cataclysms, from the other side. But in the modern world it became one of the remedies for restoring Church from ruins in the secularised European society. Naturally, the image of St. Francis was renewed in post-modern epoch.

It is significant that the polemics about Aristotelian text concerning the comedy and therefore the medieval understanding of laughter is the central point of Eco’s novel. Some details of this novel are very important in our context. First of all, Eco involves the reader in the conflict within the Franciscan order on the issue of poverty. He chooses Franciscan William as his main protagonist. We can also mention that William is a devotee and a student of Franciscan Occam. Francis in interpretation of William looks an apologist of
laughter in Christian culture. This question was never discussed in Franciscan writings but for the post-modem it could be appropriate plot.

In the contemporary cultural and information space St. Francis has left the shadow. He even became the virtual hero of Internet and personage of the mass culture. This process was intensified in 1960th years after academic studies and theological papers looked otherwise at the problems of play, carnival. It prepared the base for consideration of Christianity from a new point of view. Harvey Cox allocates the crucial role in revival of vital elements in culture to Franciscans. But the prototype is Christ. ‘Enter Christ the harlequin: the personification of festivity and fantasy in any age that had almost lost both. Coming now in greasepaint and halo, this Christ is able to touch our jaded modern consciousness, as other images of Christ cannot’ (Cox, p. 209).

This appeal was heard. Many monographs and articles of the last several decades have considered Bible as a whole or its separate parts from the point of view of the joy and laughter's presence. These publications outline trend of the modern theology that can be defined as theology of laughter. Two processes influenced modern consciousness and cultural forms. These are carnivalization of Christianity and Christian apology of carnival. St. Francis could be applauded in both contexts. But in the so-called Theology of joy the concept of joy could be seem to be inspired by his personality as well as tragic experience of a western man.

In the Middle Ages laughter was practical phenomenon but it was mainly condemned in monastic rules and ecclesiastical writings. In post-modern it became a theological topic. The basic thesis of theology of laughter is to not shut eyes to evil, absurdity and injustice, but to see all this as aspects of the Divine comedy. Theology of play [Moltmann, D.L. Miller] can be considered as one more precondition of theology of laughter. Medieval historian Heizinga elaborated it and M. Bakhtin developed its potential in the concept of a carnival and grotesque. It is reasonable to note here that Derrida used concept of play to interpret consequences of death of the God. This process was accompanied by growing interest in the play and carnival in the 20th century fiction. The titles of the novels Comedians by Graham Green and Antic Hay by Aldous Huxley became symbolic for European culture because they embody perception of life as a grotesque carnival or a sombre performance, where there are no positive

---

heroes, but the so-called anti-hero dominates. In the post-modern all the complex of theological concepts and images was involved in this play, including Christ himself. His image has undergone transformations towards conscious carnivalization or even the acquiring of tragicomic, farce, grotesque features. Such metamorphoses reflect general secularisation of the culture, modification and liberalization of Catholic attitude to some forms of human spiritual activity, with the ironical outlook of the modern intellectuals, strengthening interest to comic in its various forms.

The fool exceeds the social borders. St. Francis exceeds the human borders in order to come nearer to divine. However, for St. Francis the divine is the image of Jesus Christ first of all, Jesus Christ in his human form and in his suffering. The jester, jurodivyj and martyr comprehend joy through suffering and marginality. They are often social outcasts. In the modern western society there are almost no boundaries, which a fool exceeds to become outcast. It is harder to become an outcast. Many phenomena, which were received as marginal, now become Bohemian and thus popular and fashionable.

So, we can conclude, St. Francis combined the features of a saint, a martyr and a fool. Every of these ones were present in the image of Jesus Christ in some way. In the first two cases, Christ himself was a model for all subsequent saints and martyrs. He had been perceived as a buffoon and a fool only by the unbelieving and ignorant consciousness until recent time. St. Francis imitates Christ, as far as this is possible, in the many important episodes of his life and sermon up to stigmata. But, in contrast to Christ, he can be imagined as smiling, joyous and laughing. He sings, dances and preaches to animals. He enriched the concept of a fool by the holy naiveté, sincerity and very personal relation to the God, preserving to a certain degree carnival nature of this personage. At the same time St. Francis felt true and perfect joy through self-humiliation and suffering. St. Bonaventure, Peter John Olivi and others Franciscan writers have developed this theme in their letters, treatises and writings. In the post-modern discourse christocentric and joyous vision of the Christianity and the Catholic saint, who embodied this vision, were actualised. Naturally the Franciscan tradition was renewed. It provoked sacralization of the concepts of carnival, play and foolishness and their moving in theological context. But on the other hand Christ has become too human and his carnivalization has become possible. It has been the result of ‘the Death of God’ and other phenomena in the so-called religious postmodernism.

St. Francis did not speak directly about the laughter and did not preach it. In the post-modern discourse it is Franciscan concept of joy that appears on the foreground. But it is the concept refracted through the post-modern and it means it has provoked the theme of laughter. In the Franciscan tradition the themes of laughter, death and marginality are combined, it is that the contemporary western culture seeks. In the Orthodox Christianity the concept of joy remained self-sufficient. Catholicism has touched the theme of death and suffering during the whole its history much more. That is why the theme of laughter as a theological theme became more organic here. Only this laughter, refracted through the irony, carnival and grotesque, was adequate to the modern and post-modern perception.
Bibliography

negative theology is a discourse concerning God that attempts to describe God by negation. Its central idea is that God is hidden and ineffable. This article tries to provide a new argument for the view that the ineffability of God presupposes both metaphysical realism and the attainability of semantics. If a contrary view obtains (thus, if anti-realism and the ineffability of semantics prevail), everything we say is inexpressible of its meaning. Thus, if semantics is unattainable (or if anti-realism is adopted), the meaning of any word is inexpressible, and the ineffability of God looses content.

‘God is Ineffable’

The basics of negative theology are nicely summed up in Augustine’s famous phrase from De trinitate, ‘before we can know what God is, we can already know what He is not’. Accordingly, the classic negative or apophatic theology forbids us to follow the natural ways of thinking and to trust in the ability of human thinking to grasp God. Therefore, it is said that God is hidden and ineffable.

Mystical theologians, such as Denys the Areopagite, Moses Maimonides, and Meister Eckhart, assumed the notion about the hiddenness of God and tried to say in the first place what God is not: God is neither this nor that; he is free of all names and bare of all forms, wholly empty and free, one and simple. However, some of their expressions do say something affirmative about God, even though negative theologians basically questioned both affirmative and

---

1De trin. VIII, ii, 15 (Augustine 1981, 246)
negative predicates of God. Eckhart argued that all positive statements about
God are improper expressions, since they posit nothing in him. Again, no
negation, nothing negative belongs to God, except the ‘negation of negation’,
for God cannot deny himself.¹ Thus, no concept can grasp God such as he is in
himself, and any human concept of God is necessarily inaccurate. Therefore the
terms we use of things must be denied of God.

Moses Maimonides, who strongly influenced Thomas Aquinas, among
others, maintained that we comprehend only the fact that God exists, not his
essence. Consequently it is a false assumption to hold that God has any positive
attribute. Denys, for his part, said, in the final lines of The Mystical Theology,
that God is beyond affirmation and negation:

It [God] cannot be spoken of and it cannot be grasped by understanding.
… There is no speaking of it, nor name nor knowledge of it. … It is
beyond assertion and denial.²

Similarly Plotinus said that nothing can be affirmed truly of God. Therefore,

We must be patient with language; we are forced for reasons of
exposition to apply to the Supreme terms which strictly are ruled out;
everywhere we must read ‘so to speak’.³

Thus, Plotinus considered the expressions about God to be incorrect if
taken literally. The God-talk is figurative and anything pertaining to God
should be spoken carefully.

Negative theology suggests a constraint on God-talk by saying that God is
ineffable. However, this is problematic because by saying that God is ineffable,
one is constrained to admit that ineffable can, in some sense, be expressed,
which is a contradiction. A way out of this problem is offered by Thomas
Aquinas who greatly contributed to the interplay of positive and negative
theology. As is known, he maintained that the God-talk is analogical, as
opposed to being univocal or equivocal. In other words, the language about
God runs between two extremes: The words do not have only one and the same
meaning nor a completely different meaning when they are used of God and
humans. Rather, the words have to some extent similar and to some extent
different meanings when they are predicated of God and man. Aquinas argued
that when God is said to be good, the meaning is that what we call goodness in
creatures preexists in God and does so in some higher mode.⁴ Thus, by
analogy, we can say there is some basis of comparison, because we have an
idea of what good is and our ideas derive ultimately from God in the first place.
In other words, Aquinas thought that the terms we use of things apply in some
higher way to God and apply to God insofar as he is the cause of all. However,

¹Commentary on Exodus n. 44 and 74 (McGinn 1986, 56-57 and 68)
²Pseudo-Dionysius 1987, 13
³Enneads VI.8.13 (quoted from Louth 1989, 38)
⁴Sth I q13 a2 (Thomas Aquinas 1964b, 55)
Aquinas also said that every term used in reference to God must in an important sense be denied, for ‘what he is not is clearer to us than what he is’.1

In recent decades, negative theology has gained renewed interest. The deconstructionists, such as Jacques Derrida and Emmanuel Lévinas, as well as the postmodern theologians Jean-Luc Marion, Michael J. Scanlon and Merold Westphal, among others, abandon the metaphysical idea of God and theology to prove the existence of God. Hence they abandon what Heidegger called onto-theology and deny the being of God. In its place, Lévinas, Derrida and Marion invoke concepts such as ‘otherwise than being’, ‘God without being’, and ‘God without being God’. Derrida also spoke about ‘religion without religion’ and ‘religion about which no one knows anything’.2 In this way, deconstructionists and postmodern thinkers borrow style and expressions of negative theology.

Based on what we know about the history of theology and philosophy, in the early church and in the Middle Ages theologians tended towards what we would call (metaphysical and semantical) realism rather than towards anti-realism. It hardly comes as a surprise that this also applies to negative theologians and to the mystics who speak about divine ineffability. Nowadays, both negative theology and ineffability are coupled with anti-realism rather than with realism. This becomes especially clear in Derrida who explicitly distinguished between his thinking and negative theology. According to him, negative theology does not finally break with the onto-theological tradition that determines western thinking. In this tradition the supreme reality is ultimately seen as being timeless, constant and unchanging. The developers of negative theology, such as Denys the Areopagite, regard God as a suprarational ultimate reality, which conveys, according to Derrida, a return to kataphatic or positive theology.3 Derrida’s own deconstructive approach, for its part, forsakes the assumption that there are original, true and ultimate meanings.

Based on negative theology and the deconstruction of God-talk, one might think that the next step would be to adopt quietism and to refrain totally from language about God. From the history of theology, we know several examples of this type of pondering. In the fifth century, the Church Father Augustine wondered whether he should say anything at all about God’s Trinity. After considering this option, he reasoned that ‘the formula three persons has been coined, not in order to give a complete explanation by means of it, but in order that we might not be obliged to remain silent’.4 The doctrine of the Trinity is thus an alternative to silence. Karl Barth, for his part, argued that because theologians are human, they cannot speak of God, but even then they ought to speak. Augustine and Barth were unwilling to accept theological quietism, which is an approach that would seem to agree with Wittgenstein’s assertion from the Tractatus: ‘Everything that can be thought at all can be thought

---

1Sth I q1 a9 (Thomas Aquinas 1964a, 35)
2Derrida 1995, 49; 2002, 97-98; Caputo & Scanlon 1999, 12, 15 and 197
3Derrida 1989
4De trin. V, ix, 10 (Augustine 1981, 188)
clearly. Everything that can be put into words can be put clearly. … What we cannot speak about we must pass over in silence.\(^1\)

Some contemporary theologians and philosophers, partly encouraged by the Tractatus thesis, conclude that meaningful talk about God is impossible. The fact that the views about God cannot be scientifically verified or falsified is a major reason why many philosophers of the early 20\(^{th}\) century considered religious language factually meaningless. The underlying basis of their conclusion was the logical positivist’s anti-metaphysical restriction of meaningful language to analytical (mathematical-logical) and empirical (verifiable, at least in principle) statements. Although the influence of logical positivism has declined, the question of the nature and function of religious language is very much alive.

**Various Meanings of ‘Ineffability’**

The idea that God is ineffable is the epitome of negative theology. As we have seen, the concept of ineffability is confusing, however, because something that can be called ineffable is not ineffable because we have already grasped something about it. Thus, by using the word ‘ineffable’, the addresser admits that what she speaks about is somewhat expressible in the language anyway. For that reason, Augustine argued that if we say of something that it is ineffable, we are actually contradicting our own statement:

What I have spoken would not have been spoken if it were unspeakable. For this reason God should not even be called unspeakable, because even when this word is spoken, something is spoken. There is a kind of conflict between words here: if what cannot be spoken is unspeakable, then it is not unspeakable, because it can actually be said to be unspeakable. It is better to evade this verbal conflict silently than to quell it disputatiously.\(^2\)

Contrary to Augustine’s advice to remain silent, it has been suggested that we should restate the claim that God is ineffable. According to the proposed restatements, it is not the case that any human concept is applicable to God, and sentences of the form ‘God is \(x\)’ (where \(x\) is any descriptive predicate) are incorrectly formed.\(^3\) These claims convey the supposition that our language cannot speak about God appropriately, and that our linguistic categorisations fail to grasp God. It has also been argued that the term ‘ineffable’ does not refer to God but to the limits of human knowing in the first place. In addition, particularly in Judaism, God is regarded as ineffable in the normative sense that one should not use his name in vain.\(^4\) Consequently, one might think that ineffability is not a predicate of God. Rather, the talk about ineffability expresses a rule of the grammar of faith that directs our way of speaking of

---

\(^1\)Tractatus 4.116 and 7 (Wittgenstein 1995, 26 and 74).


\(^3\)Alston 1956, 508-510; Gutting 1982, 57.

\(^4\)Ex. 20:7.
God. The rule states that we should use the word ‘God’ carefully and sparingly, and that, if anything, we should say what God is not. This idea is inspired by numerous passages in the Old Testament that speak of the inscrutability and incomprehensibility of God.¹

The words ‘ineffable’, ‘indefinable’ and ‘indescribable’ are near-synonyms. Similar expressions are ‘beyond words’ and ‘no words can convey’. Although God is said to be ineffable, many theologians are reluctant to say that God is totally indefinable and indescribable. This is understandable as, if God were totally indefinable and indescribable, ‘God’ would be a meaningless string of letters and theological statements would be baseless and void. In addition, one might ask what point there is to believe in God, or even to use the word ‘God’ if its meaning is totally unknown even to theologians who supposedly are experts of the God-talk.

Ninian Smart challenged the view that the ineffability of God arises from a lack of adequate expression. He pointed out that we often use the expression ‘no words can convey’ in order to convey something, for instance our deep gratitude to someone. To speak about the lack of words is thus a way to express something. Equally, to say that God is ineffable does not indicate that we can say nothing about God. Rather, it suggests that our words are inadequate to reach the right tone and intensity when used of God. Smart argued for this view by considering feeling and emotional vocabulary. Pains are more or less intense and can be expressed appropriately by more or less intense expressions, such as ‘a severe pain,’ ‘a horrible pain’ and ‘an unbearable pain’. However, conventions are completely inadequate in relation to a really horrible pain. This is where a pain of great intensity can properly be expressed by the word ‘indescribable’. Likewise, the beauty of nature, music, or a painting can be so great, or subtle that it is expressed by saying ‘beyond words’.

It is in a similar sense that God can be said to be ineffable. In place of speaking, a person can keep silent. Silence can even be a nobler and more effective expression of worship than any word, for expressive language easily becomes devalued. One protection against semantic inflation is to retain a mysterious, rather unintelligible language, and silence. One of the reasons for sacred language and silence in religion is precisely their power to remain beyond the cliché and trite expression.²

From a logical point of view, something can be ineffable in two senses: First, if something has a form but is devoid of content, then there is of course very little to be said of the content: basically we can only say that content is lacking. It is initially in this nihilistic sense that something can be ineffable, and one could claim that secularization will just lead to this kind of ineffability: If a person is unfamiliar with religion, the process of which has been claimed to be a result of secularization, he or she may easily find religious expressions empty and meaningless. Second, something can be ineffable in the sense that language is not rich and expressive enough a medium to convey the desired content. Expressed differently, words may be insufficient and inadequate to

²Smart 1972, 28-29.
express something which, however, has content. Many writers, poets and other artists have had the experience that they have been unable to find a proper expression for what they would have wanted to express. I suppose it is in this latter sense in which different religions have often spoken of ineffability.

The word ‘ineffable’ has an additional meaning that has stimulated the discussion of the meaningfulness of religious language. In some contexts, ineffable means inconceivable or unthinkable. For this reason some philosophers have regarded the idea that religion concerns itself with the ineffable as an indication of the factual meaninglessness of religious language. This stance is quite understandable, for how could we think of something which cannot be spoken and expressed in the language? If something is ineffable, does it follow that we cannot describe, define nor even think of it with words? And if, nonetheless, we can think of something that is ineffable, must that be a kind of nonlinguistic thinking?

It has been argued that unthinkableness implies impossibility. Thus, if something is unthinkable, it is logically impossible. Examples would be a round quadrangle, or a pencil that is not a pencil. Some opinions add that ineffability also implies impossibility. Thus, if something is ineffable, it makes no sense, or its concept is contradictory, unintelligible and meaningless. However, when God is said to be ineffable, it has not traditionally been understood to imply that the concept of God is incoherent or meaningless. Rather, a long tradition of negative theology since Philo and Middle Platonism has argued that God is unknowable in himself, and that the essence of God cannot be encompassed by human conceptions. God is thus beyond knowledge and beyond the reach of the powers of human understanding. No concept can grasp God such as he is in himself. It is in this sense that God is understood as being ineffable. Hence it would be more accurate to say that the concept of God is obscure than to say that the concept of God is meaningless.

If something is ineffable, it would seem to entail that we cannot properly know the concept that is related to it, and if we do not know a concept, neither do we know whether that concept is coherent or contradictory. One could also put it this way: If God is ineffable, we cannot know the logic of the concept of God. Thus, we do not know what inferences we can make from the sentences about God, nor do we know in what way the sentences about God relate to each other. The question of what inferences we can make from the sentences about the Trinity was much discussed by medieval theologians. They concluded that even if the concept of God is unknown to natural reason, it does not imply that the concept is impossible.

I would like to put forward a thesis that there are two main views of ineffability. The first can be called the neo-Platonic view. According to this view, ineffability refers, in the first place, to that which is indefinable and incomprehensible. In other words, if something is ineffable, it is, according to the first view, beyond the possibilities of language and human understanding. The second position, for its part, can be called the Kantian view. According to this, ineffability refers to that which is unknowable. In other words, if something is ineffable, it transcends or is beyond the limits of all possible experience and knowledge. I am not ready to say that these two views are
autonomous and distinct. Rather, they are abstractions and appear to be intertwined.

Semantics is Expressible

So far we have discussed various meanings the term ‘ineffability’ can have, and we have seen that a variety of philosophical questions and problems are involved in the concept of ineffability. When we now move to consider the ineffability of semantics, one of the meanings of ‘ineffability’ is particularly relevant: inexpressibility.

In brief, there are two major views of semantics: Either we can define the way in which words and sentences relate to reality, or we cannot; either we can express the relations of meaning in the language, or we are unable to do that. These alternative views are related to the much-discussed distinction between metaphysical realism and anti-realism.¹ The realist idea is that external objects exist independently of the mind and that our words enter into uniquely determinate relationships with mind-independent objects. The anti-realist view is that the terms of our language are always related to their referent by other terms or concepts. The adherents of this latter view say that reference is mediated by a conceptual structure. Anti-realists also claim that there is no fact of the matter about what the expressions of language refer to. For this reason, semantics is inscrutable. Indeterminacy of reference and the inexpressibility of semantics are thus related facts. According to the anti-realist view, semantics is inexpressible just because there is nothing to be expressed.

Viewed from the tradition, negative theology and ineffability are allies of realism rather than of anti-realism. Accordingly, the traditional theological interpretation of ineffability seems to presuppose the notion that, in general terms, semantics is attainable. If a contrary notion obtains, everything we say is inexpressible of its meaning. Consequently, the notion of the ineffability of God would suffer a semantic inflation and would be redundant: since all meanings are inexpressible it is pointless to say that God is ineffable. Thus, if semantics is unattainable (or if anti-realism is adopted), the meaning of any word is inexpressible, and the ineffability of God looses content. For this reason, the ineffability of God presupposes both metaphysical realism and the attainability of semantics.

Many philosophers (Wittgenstein, among others) have held that we cannot speak of the relation that language bears to the world. We can just say that, in virtue of that relation, words and sentences of our language are meaningful. Wittgenstein argued that to investigate the conditions on which our statements relate to reality, we should be able to observe the relation between language and the world from the viewpoint that is not already determined by language and its meanings. To attain such a language-independent viewpoint to semantics is impossible, because the meaning relations are always dependent on language. In a much-quoted passage from *Tractatus*, Wittgenstein said,

'What finds its reflection in language, language cannot represent. What expresses itself in language, we cannot express by means of language. ... What can be shown, cannot be said'.1 In other words, it is not possible to check whether our statements correspond to reality, because it would require the renouncing of conceptual practices, which we cannot renounce without losing our concepts at the same time. Thus, it is not possible to compare a sentence and a fact, and thus semantics is inexpressible.

Jaakko Hintikka (in accordance with Hans-Georg Gadamer and Jean Van Heijenoort) has labelled the above notion as ‘language as the universal medium’.2 Language is universal, because it cannot be escaped, and the attempts to examine language by simultaneously not using language are doomed to fail: language cannot be seen outside, and we cannot get behind language. Since the relation between language and reality is inexpressible, we have to limit our semantic investigations to the relations that linguistic expressions bear to other linguistic expressions. We cannot say anything about the relation that signs bear to the world. Consequently, the world as such remains unknown to us.

Many people seem to be ready to accept the notion of the universality of language, if it is presented as above. According to that notion, it is impossible to express the conditions of correspondence between our assertions and reality. However, the rival notion better corresponds to our everyday intuition of language and meaning. This other view, which Hintikka calls ‘language as a calculus’, considers possible all that which the universalist holds to be impossible. The notion of language as a calculus asserts that we are not bound to our language. Rather, language is our servant and we are its master, because we can determine the ways in which language functions.

In Hintikka’s view, the inexpressibility thesis is based on the confusion in the concept of semantics itself. When it is said that the meaning of language cannot be expressed because its expression always requires language, it is assumed that the whole net of meanings should be expressed in one go. In reality, we can shed light on the meaning of language part by part by using one part of language as the light and other part as what is lit. Semantical relationships are thus inexhaustible rather than inexpressible. The same is the case with the extensions of language towards new contexts of use. We can always reinterpret our words and sentences, and we can attach new meanings to our expressions. So, the conditions of correspondence between our statements and reality can be changed. However, this is neither an arbitrary operation nor one restricted to a particular person, because for words to have a meaning, a relevant context must be invoked in which they have an application that can, at least in principle, be known also by others than their inventors.

Thus, words and sentences are meaningful in virtue of the interpretations associated with them, and we do not need to confine our interpretations to actual or established models, but can also imagine and construct new possible interpretations. Consequently, semantics is inexhaustible, for language can repeatedly be reinterpreted, and the meaning relations can systematically be

---

1 Tractatus 4.121 and 4.1212 (Wittgenstein 1995, 26).
changed. Moreover, the meaning relations are expressible, and we can systematically investigate the conditions on which our statements correspond to reality. That is to say, we can express the meanings of our sentences. This takes place by telling what ideas, presuppositions, perceptions, practices, or contextual settings our sentences are based on. We can also elucidate the meaning of our sentences by explaining what inferences we can make from them. All these procedures are in accordance with our everyday use of language.

Conclusion

The two notions of language presented here—‘language as the universal medium’ and ‘language as a calculus’—offer different interpretations of the ineffability of God. According to the thesis of the universality of language, the semantic relations are inexpressible, and equally we cannot express the conditions of correspondence between reality and the utterances about God. In other words, we cannot express the way in which God-talk relates to reality. Thus, the relation that the utterances about God bear to reality is unattainable, and we have to limit our semantic investigations concerning religious language to the relations between linguistic expressions only. However, this inexpressibility of semantics concerns just any sentence and not merely religious language. There is no reason to believe (and every reason to deny) that the tradition of negative theology has considered the ineffability of God as a mere exemplar of the general unattainability of semantics.

The notion of language as a calculus maintains that the semantic relations are definable. The definition takes place in a metalanguage, taking a fragment of natural language as metalanguage. Thus, there is no universal barrier related to the nature of language, due to which we would, in turn, be unable to investigate the way in which language is directed to reality. We can assume that this concerns religious language as well. There is thus no universal barrier which would prevent us from defining religious concepts such as the concept of God. Language as such does not hide God from humans. On the contrary, the idea humans have in the language is an idea of a hidden God.¹ The notion that God is ineffable is a pragmatic rule for the God-talk. This rule advises abstention from theological speculations, and suggests that we should say what God is not, rather than what he is.

Bibliography


¹Phillips 2000, 81.


The Franciscan Concepts of Joy, Laughter and Foolishness in the Post-modern Context
The traditional problem of evil can be formulated in the following way:

“If God exists, how can evil be explained? For an omnipotent being would have the power to prevent any and all evils if wanted to; an omniscient being would know all about them; and a perfectly good being would want to prevent/eliminate all the evils it could. Thus, it seems, if God existed, and were omnipotent, omniscient and perfectly good, there would be no evils."

On closer examination, however, it becomes clear that the conditional of the above mentioned *modus ponens* is a *non sequitur*[^3]: The conditional claim would be false without the additional assumption that God is rational, that is, “being based on, or agreeable to, reason”[^4], which relies on the law of bivalence (that every proposition takes exactly one of two truth values, namely true or false) and the laws of identity, non-contradiction and of the excluded middle. Therefore, the correct *modus ponens* of the problem of evil would go as follows:

[^1]: The basic idea for the following considerations hark back to fruitful discussions with Peter Prechtl. This essay is dedicated to his memory.
[^3]: I am thankful to Dimitris Platchias for his clarifying criticism of an earlier version of the introduction.
Deconstructing God: On the Logical and Conceptual Clarification of the Problem of Evil

Premise one: If God exists and is omnipotent, omniscient, perfectly good and rational, then there is no evil.
Premise two: God exists and is omnipotent, omniscient, perfectly good and rational.
Conclusion: Therefore, there is no evil1.

Deconstructing God: Against the Supplement of Rationality

In order to become more clear about the problem of evil, it is necessary to examine the unspoken assumption in premise one of the modus ponens in more detail for its validity. For this, I would like to start from the beginning, that is, from epistemology. Following Niklas Luhmann, we can consider knowledge as a distinguishing operation that indicates something:

One can answer the question ‘How is knowledge possible?’ with ‘By the introduction of a distinction’.1

Therefore, the act of cognition consists in denoting something:

“The activity of observing establishes a distinction in a space that remains unmarked, the space from which the observer executes the distinction. The observer must employ a distinction in order to generate the difference between unmarked and marked space, and between himself and what he indicates. The whole point of this distinction (its intention) is to mark something as distinct from something else.”2

Hence, “every observation […] uses a distinction to mark one side (but not the other)”, which leads directly to the law of the excluded middle (A or not-A) as formulated by Aristotle:

[…] there cannot be an intermediate between contradictories, but of one subject we must either affirm or deny any one predicate. This is clear, in the first place, if we define what the true and the false are. To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true; so that he who says of anything that it is, or that it is not, will say either what is true or

1Basically, the same structure goes also for the so-called „Problem of Providence“, which asks how human beings can be free if God is omniscient and therefore knows the outcome of every human decision before it is taken:
Premise 1: If God exists and is omniscient, then he knows the outcome of every human decision before it is taken.
Premise 2: God exists and is omniscient.
Conclusion: Therefore, there is no human freedom of will.
2Niklas Luhmann: The cognitive program of constructivism and the reality that remains unknown, p. 130.
3Niklas Luhmann: Art as a social system, p. 54f.
what is false; but neither what is nor what is not is said to be or not to be – Tertium non datur.

Besides, a distinction implies that one cannot be on both sides at the same time, which is expressed in the law of non-contradiction (not (A and not-A)):

[…] the most indisputable of all beliefs is that contradictory statements are not at the same time true. […] Now since it is impossible that contradictories should be at the same time true of the same thing, obviously contraries also cannot belong at the same time to the same thing.

Nevertheless, one is able to change the perspective and observe from the other side of the distinction, but only after that. For the passing of the boundary, called “crossing” by George Spencer Brown, takes time. Being the one, who distinguishes, one can have an eye on that, what has been differentiated from the other; though, while doing this, the own distinction cannot be seen:

The impossibility of distinguishing the distinction that one distinguishes with is an unavoidable precondition of cognition. The question of whether a given choice of distinction suits one’s latent interests arises only on the level of second-order observation.

Following Heinz von Förster, Luhmann calls this “second-order observation”, which, in contrast to first-order observation, is an observation of an observation, also called “second order cybernetics”. For him, this concept is of great importance for his (epistemological) constructivism:

The usual understanding of the observations of observation focuses above all on what an observer observes (distinguishing thereby between subject and object, but concentrating above all on the object). Constructivism describes an observation of observation that concentrates on how the observed observer observes. This constructivist turn makes possible a qualitative change, a radical transformation, in the style of recursive observation, since by this means one can also observe what and how an observed observer is unable to observe. In this case

---

3Metaphysics 1011b.
4Aristotle: Metaphysics 1011b.
5With regard to Luhmann’s explanations, see, in general, George Spencer Brown: Laws of form.
6Niklas Luhmann: The cognitive program of constructivism and the reality that remains unknown, p. 141.
7First-order observation operates on the level of the factual:
   “First-order observation is an indication of something in opposition to everything that is not indicated. In this kind of observation, the distinction between distinction and indication is not thematized. The gaze remains fixed on the object” (Niklas Luhmann: Art as a social system, p. 61).
Deconstructing God:  
*On the Logical and Conceptual Clarification of the Problem of Evil*

one is interested in his blind spot, that is, the means by which things become visible or nonvisible.¹

For Luhmann, the so-called “blind spot” of observation is the distinction, that is necessarily drawn by observing something; hence, Luhmann points out:

As we know from operative epistemology, [...] the activity of observing occurs in the world and can be observed in turn. It presupposes the drawing of a boundary across which the observer can observe something (or himself as an other), and it accounts for the incompleteness of observations by virtue of the fact that the act of observing, along with the difference of the observation that constitutes it, escapes observation. Observation therefore relies on a blind spot that enables it to see something (but not everything). A world equipped to observe itself withdraws into unobservability. Or, to use a more traditional formulation, the unobservability of the observing operation is the transcendental condition of its possibility.²

And yet, it is possible to take a glance at this distinction; then, the original distinction occurs in the distinguished, called “reentry” by Luhmann as William Rasch explains:

One makes a distinction and then operates on one side of the distinction that is made. One issues diatribes against irrationality, for instance, from within the realm of rationality. In like manner, one consciously discusses the unconscious. Or again, one communicates about the incommunicable. Such possibilities are enhanced when the distinction – communicable/incommunicable, say – reenters the space from which communication proceeds. One calls the incommunicable silence, or the sublime, and thus turns silence into a form of communication.³

With regard to epistemology, the distinction between self-reference and external reference, i.e. the difference between internal and external states in general, makes up “the condition of possibility for observation”⁴ at all:

Operational closure seems to be the necessary *empirical* condition of observations. Without closure, the system would continually mix up its own operations with those of its environment, conscious states with external states or words with things. It could not make the (reentering) distinction between self-reference and external reference. It could not

---

¹Niklas Luhmann: The cognitive program of constructivism and the reality that remains unknown, p. 140.  
²Niklas Luhmann: Art as a social system, p. 56f.  
⁴Niklas Luhmann: Art as a social system, p. 57.
even match external and internal states. It could not separate the observer from the observed. It could not produce cognition.5

According to Luhmann each distinction or difference establishes a so-called “form”, consisting of what was distinguished (the inside of the form) and all the other (the outside of the form):

A form is thus something with two sides that are distinguished. (…) A form has two sides; (…) A form is established by fixing a boundary. This leads to the separation of two sides such that getting from one side to the other requires another operation to cross the boundary. Establishing a form is thus distinguishing.6

Therefore, any form exists only as “the form of the form” via second-order observation; as on the level of first order observation, the observer does not know the distinction, that makes possible his observation at all. As theory-speakers, speaking about the phenomenon of form, however, we are always-already second-order observers: We observe observers. Hence, by observing a form, i.e. a distinction of two sides, a form (of a first-order observation) occurs in a form (of a second-order observation) – a reentry takes place:

A distinction marks a space and then is inserted into the distinguished again. Then, it occurs twice: as starting distinction and as distinction in the distinguished. It is the same and not the same. It is the same as the point of reentry just consists in applying the same distinction to itself recursively; it is not the same as it occurs in another already distinguished space.1

In this respect, cognition stands out due to its paradoxical structure of revelation and concealment: Observation is only possible because of the concealment of the distinction revealing reality; however, Luhmann points out that “[s]uch a paradox is not simply a logical contradiction (A is non-A) but a foundational statement: The world is observable because it is unobservable. Nothing can be observed (not even the ‘nothing’) without drawing a distinction, but this operation remains indistinguishable. It can be distinguished, but only by another [and not the same] operation. […] Obviously, this makes no sense. It makes meaning. It makes no common sense; it uses the meaning of ‘para-doxon’ to transgress the boundaries of common sense to reflect what it means to use meaning as a medium.”2

With regard to these considerations, God – per definitionem being absolute (and not relative) – cannot be thought of as being bound by any distinctions such as strong or weak, male or female, etc. Instead, God is to be regarded as the coincidence of opposites or coincidentia oppositorum as Nicholas of Cusa stresses:

---
5Niklas Luhmann: Deconstruction as second-order observing, p. 106.
6Niklas Luhmann: The paradox of form, p. 16.
2Niklas Luhmann: The paradox of observing systems, p. 87f.
Deconstructing God:
On the Logical and Conceptual Clarification of the Problem of Evil

[...] [by comparison] Absolute Oneness, to which nothing is opposed [cui nihil opponitur], is Absolute Maximality, which is the Blessed God.³

In this respect, God resembles the paradoxical middle or Tertium of classical, Aristotelean logic (grounded in the laws of identity, of bivalence, of the excluded middle and of non-contradiction) as Nicholas of Cusa explains:

For in the case of God we must, as far as possible, precede contradictories and embrace them in a simple concept. For example, in God we must not conceive of distinction and indistinction as two contradictories but [must conceive of] them as antecedently present in their own most simple Beginning, where distinction is not anything other than indistinction.⁴

Consequently, our concept of God is essentially paradoxical according to Nicholas of Cusa: In God, we have to think of distinction as indistinction or difference as unity respectively⁵. This, however, results in a paradox as Niklas Luhmann stresses: “Using both sides [of a distinction; H.M.] at the same time would infringe upon the purpose of the distinction. It is not possible; it would result in a paradox since one would have to call at once what is different the same.”¹ In other words: “Posing the question of unity would lead to the paradox of the identity of what is different.”²

Thought, to speak with Johann Gottlieb Fichte, “eliminates itself”³ by thinking God. In this respect, God is “the ultimate paradox of thought: to want to discover something that thought itself cannot think”⁴; for our discursive thinking and knowing always already makes use of binary distinctions governed by the laws of the excluded middle and of non-contradiction amongst others as made clear above with reference to Aristotle. In contrast to Aristotle, however, Nicholas of Cusa shows that we have to look through the glasses or beryl of the coincidence of opposites in order to grasp the truth about God:

³De docta ignorantia I, 5 n.14.
⁴De docta ignorantia I, 19 n.57.Similarly, Immanuel Kant says:
“For what necessarily forces us to transcend the limits of experience and of all appearances is the unconditioned, which reason, by necessity and by right, demands in things in themselves, as required to complete the series of conditions. If, then, on the supposition that our empirical knowledge conforms to objects as things in themselves, we find that the unconditioned cannot be thought without contradiction” (Critique of pure reason, B xx).
⁵In this context, one might think of Heraclitus: „God is day and night, winter and summer, war and piece, surfeit and hunger [...]“ (Fr. 36; translated by John Burnet: Early Greek philosophy, p. 149).
¹Niklas Luhmann: The paradox of form, p. 18.
²Niklas Luhmann: The paradox of form, p. 25.
⁴Sören Kierkegaard: Philosophical fragments, p. 37.
In harmonizing all the other philosophers, Aristotle said that the beginnings which are present in a substance are contraries. And he named three beginnings: matter, form, and privation. Although more than all the other philosophers Aristotle is held to be the most careful and most acute reasoner, I think that he and all the others utterly failed in regard to one point. For since the beginnings are contraries, those philosophers failed to arrive at a correct understanding of that third, assuredly necessary, beginning [viz., privation]. This failure occurred because they did not believe it to be possible that contraries coincide in that beginning, since contraries expel one another. Hence, from a consideration of that first principle which denies that contradictories can both be true at the same time, the Philosopher showed that, likewise, contraries cannot be present together. Our beryl makes us see more acutely, so that, in the Uniting Beginning, we see opposites prior to duality, i.e., before they are two contradictories. [The situation is] as if we were to see the smallest of contraries coincide (e.g., minimal heat and minimal cold; minimal slowness and minimal fastness; etc.), so that they are one beginning prior to the duality of both contraries.

[...] But if Aristotle had understood the beginning which he calls *privation* – understood it in such a way that privation is a beginning that posits a coincidence of contraries and that, therefore, (being ‘deprived’, as it were, of every contrariety), precedes duality, which is necessary in the case of contraries – then he would have seen correctly. But out of fear of admitting that contraries are present at the same time in the same thing, he shunned the truth regarding this beginning.1

Due to our binary thinking (or *discursus rationis* as Nicholas of Cusa holds it), God absconds again and again from our attempts to conceptualize him, which negative theology demonstrated accurately3 – as a poet-philosopher once remarked: “We seek everywhere the unconditioned [das Unbedingte] and always find merely conditions [Dinge].”4

Therefore, our concept of God is essentially paradoxical and the assumption that God is rational serves as a mere *supplement*5 to cover this conceptual inconsistency or incompleteness. This fundamental circumstance

---

1Nicholas of Cusa: De beryllio, 26.
2Nicholas of Cusa: De beryllio, 22.
3See, for example, Pseudo-Dionysius the Aeropagite: The mystical theology, p. 141.
4Novalis: Miscellaneous Remarks, p. 383. Cf. also John Caputo: “God is always and everywhere, in all the epochs, essentially withdrawn from the world, even as faith says He is omnipresent. His very self-giving is self-withdrawing, a-letic. God is never *given* in some sheer excess of presence or plentitude […]. His presence is deferred even as it is revealed […]. God’s presence is always caught up in the play of presence and absence” (Radical hermeneutics, p. 279f.).
5Being a term of Jacques Derrida’s thinking, “the supplement adds itself, it is a surplus, a plentitude enriching another plentitude, the fullest measure or presence. It cumulates and accumulates presence” (Of Grammatology, p. 144). In the case of the problem of evil, the assumption of God being rational serves as a supplement, because it pretends to make God (and his intentions) transparent or intelligible to us as Susan Neiman points out: “The problem of evil began by trying to penetrate God’s intentions” (Evil in modern thought, p. 281).
becomes manifest in several intractable paradoxes, that are closely connected with the concept of God in the philosophical and theological tradition: The Paradox of Omnipotence\(^1\), the Paradox of Providence\(^2\) and (with regard to Christianity) the Paradox of Incarnation\(^3\). The most illustrative one, however, is the Paradox of Omniscience\(^4\) that takes reference to Georg Cantor’s theorem\(^5\) that every set \(z\) will be smaller than its power set \(x: z < \{x \mid x \subseteq z\}\), as Patrick Grim points out:

> The argument will hold […] for any system interpreted as applying to a domain of objects – those things the system is taken to be about – and including a range of predicates applicable within the system. If any such system is self-reflective – if each of its predicates can also be taken as an object to which the system applies – it will have at least as many objects as predicates. But taking properties purely extensionally, by Cantor’s power set theorem there will be more properties of individual objects of the system than corresponding predicates with which to express them. Some genuine property of some object of the system – and thus some truth – will be inexpressible, and any such system will prove expressively incomplete.\(^6\)

As any omniscient mind hence “would surely be self-reflective in at least the following sense: among its objects of knowledge – among those things it knows something about – would be its own conceptions of properties (…), it will have at least as many objects of knowledge as conceptions of properties, since each of the latter is also an object of knowledge. But by Cantor’s argument there will be more actual properties of its objects than objects themselves. Actual properties will outnumber its conceptions of properties, and

---

\(^1\)If God is almighty, then he can *per definitionem* create a stone which he cannot lift, which means that he is not almighty. For this and the other given paradoxes as well as for further examples see The Oxford companion to Christian thought, p. 515f.

\(^2\)See above, p. 1, fn.5.

\(^3\)“Eternity appears in time, immensity in measurement, the Creator in the creature […] the unfigurable in the figure, the unrarratable in discourse, the inexplicable in speech, the uncircumscribable in the place, the invisible in vision” (Bernadino of Siena as quoted by Georges Didi-Huberman: Fra Angelico, p. 35).

\(^4\)In the following, “omniscient” has the traditional meaning of “knowing all true propositions” as stated in Edward R. Wierenga: The nature of god, p. 59.


\(^6\)Patrick Grim: Logic and limits of knowledge and truth, p. 350. With regard to the notion “system”, Grim adds:

> “Note also how very thin a notion of ‘system’ is in fact required in the basic argument above. Nothing has been said to indicate that any system at issue must be formal or axiomatic, that it must generate theorems by means of demonstrations or even that it must contain a category of assertions or asserted theses. All the argument requires, in fact, is a system of expression.”
thus some genuine property of its objects of knowledge – and so some truth – will remain *inconceivable* for such a being.”¹

Conclusion

In sum, we can conclude that the necessary additional assumption in the conditional of the *modus ponens*, that God is rational, i.e. relies on the law of bivalence and the laws of identity, non-contradiction and of the excluded middle, and which originates the problem of evil as a philosophical problem, is untenable. Referring to Nicholas of Cusa, we could show that in God – being absolute – all distinctions and differences coincide and that, in this respect, God is to be thought of as *epekeina tes ousias*, as Plato already emphasized². Hence, our concept of God is essentially paradoxical. This, however, must not be understood as “the perplexity dead end of a false path but [as] every path’s point of origin”³ as we have already seen with regard to Luhmann’s explanations about cognition earlier. Rather, it should be regarded as a “veridical, or truth-telling paradox”⁴ in the sense of Quine that appears *necessarily* by grasping God “per discretionem”⁵ (as Nicholas of Cusa would say), i.e. by means of our discursive thinking and speaking. In this respect, we have to deconstruct⁶ the concept of God, which underlies the problem of evil, in order to reconstruct philosophy of religion.

Bibliography


¹Patrick Grim: Logic and limits of knowledge and truth, p. 354. For a more detailed justification, see Patrick Grim: Logic and limits of knowledge and truth, p. 347ff. Interestingly enough, the above-mentioned paradoxes (of omnipotence, of omniscience and of incarnation in Christianity) are all *paradoxes of self-reference*: In each case, one of God’s properties is applied to God himself.
²Politeia 509b. See also Meister Eckhart: Quaesitones Parisienses, p. 47,14f.
⁴Willard van Orman Quine: The ways of paradox, p. 3.
⁵Nicholas of Cusa: Idiota de sapientia I, 5.
⁶Thus, negative theology stressing that God transcends the phenomenal world “is a form of deconstruction” itself as Kevin Hart shows: “If God is understood as transcending the phenomenal world, one cannot hope to describe Him because language is restricted in its scope to the realm of the phenomenal. Similarly, if *différance* enables concepts to emerge it cannot be described adequately by concepts” (The trespass of the sign, p. 186).
−: Deconstruction as second-order observing. In: Niklas Luhmann: Theories of distinction, pp. 94-112.
−: The cognitive program of constructivism and the reality that remains unknown. In: Niklas Luhmann: Theories of distinction, pp. 128-152.
Deconstructing God: on the Logical and Conceptual Clarification of the Problem of Evil
Freud’s well-known theory of the id, ego, and super-ego was not developed by him until late in his writings in his *The Ego and the Id* and *Lectures on Psychoanalysis* in 1923 and commented upon with respect to morality and religion in his *New Introductory Lectures in Psychoanalysis*, 1933. He affirms in the latter book that the formation of the super-ego is the origin of conscience. He summarizes a famous statement by Kant about two things filling the mind with ever new and increasing admiration and awe, the starry heavens above us and the moral law within. The point of Kant was that just as we find in the heavens the universal and physically necessary laws of physics exemplified in the stars, so also do we find in our conscience the universal and morally necessary laws of morality. These two experiences of the starry heavens and the moral law are the best argument for the greatness of God, Freud affirms.¹ In a challenging comment, he grants the point about the stars but questions whether God has done such a good job of instilling the moral law within us. Many people have only a limited sense of conscientious responsibility to others or even almost none at all. Whereas Kant has tried to argue from the existence of his absolute morality, founded in the categorical imperative, shared by all people, to the necessary presuppositions of the existence of an immortal soul and of God as the final judge of human moral endeavors, Freud does not see humans as possessing this sense of conscience universally since there obvious exceptions such as the sociopath, and he argues that it cannot be basis of a solid, rational argument for God as the necessary presupposition of morality. In contrast with Kant’s essentialist argument for the innateness of the moral law rooted in pure practical reason, Freud offers a developmental theory of morality rooted deeply in both emotion

---

and reason. Freud sees morality as non-universal in character, with the possibility that deviations may arise in morality, such as in obsessive neuroses or psychoses.

Despite his criticism of Kant, Freud does find a deep psychological but fragmentary truth in "the assertion that conscience is of divine origin!"1 The id in the form of instinctual energy, including both the sexual instinct and the death instinct, is present at the earliest strivings of the infant, and the littlest child has no restrictions within itself that would prohibit any form of the quest for pleasure. Only when the parents have begun to control the child’s behavior through affectionate rewards and through punishments which both signify the withdrawal of love and the inherent pain of the punishments themselves, then the child develops psychological anxiety of the gaining and losing of the parents’ love. This psychological anxiety becomes moral anxiety when “the external restrictions are introjected, so that the super-ego takes the place of the parental function,” creating an internal control that works just as the external rewards and punishments of the parents worked to control the child.2

Even when writing in 1933, Freud states that he has not fully understood the process which creates the super-ego in the child but offers some indications of how identification of the child with the parent is involved. In identification, “one ego becomes like another, . . . which results in the first ego behaving itself in certain respects in the same way as the second.”3 Freud compares this psychological identification with oral cannibalism. For example, a member of one tribe will eat the heart of his brave enemy in order to identify with the courage of that slain enemy. There is a distinction between identification and object choice:

[W]hen a boy identifies himself with his father, he wants to be like his father; when he makes him the object of his choice, he wants to have him, to possess him; in the first case his ego is altered on the model of his father, in the second case that is not necessary.4

For Freud, ‘identification’ is a sexually invested process, whereby the child expropriates the image of the Father as a permanent, internal function of control: the boy ingests the Father to make the Father part of himself. The non-sexual process that Freud here alludes to is the process of ‘sublimation’, where non-sexual energies are invested in an object that remains external to and other than the child, thus entailing a separation and desire for the sublimated object that motivates the child to seek the chosen object.5 This sexually invested process of identification with the parent is intensified in the Oedipus complex, for example, in which the little boy suffers who has made an object-choice of the mother for his sexual longing and who finds that he must identify with and

---

become the Father in order that he come to possess his mother. ¹ He finds, however, that his identification with the Father becomes too intense, that the Father establishes himself as the continual prohibition of his desire. In this way, in what Freud calls the castration complex, the child shifts his internal identification with the Father from an abstract, sexual ideal to an entire fabric of control in the moral order of the super-ego. Successful overcoming of the Oedipus complex is required for the super-ego to reach its true strength. As the super-ego develops, it will also take on the ideals, commands, and prohibitions of others who function in a similar role to that of the parents, for example, with teachers and coaches. So the super-ego transforms, becoming more generalized and distant from one’s parents. ²

The super-ego functions in three ways in Freud’s analysis. First, it is the source of self-observation in that one can see one’s actions from the viewpoint of the other which has been incorporated in the identification of the person with the other. So a person looks upon oneself from the assumed viewpoint of the other. Second, the super-ego functions to control the behavior of the child from the incorporated parents’ prohibitions. Third, the super-ego functions to become an impulse towards perfection and the better things for which people in that culture strive. ³ In this last function the super-ego is:

... the vehicle of the ego-ideal, by which the ego measures itself, towards which it strives, and whose demands for ever-increasing perfection it is always striving to fulfil. No doubt this ego-ideal is a precipititation of the old idea of the parents, an expression of the admiration which the child felt for the perfection which it one time ascribed to them. ⁴

**Freud’s Concept of God**

What Freud has expressed here, especially in the third function of the super-ego as the vehicle of the ego ideal combined with the child’s mistaken perception of the parents as perfect role models to be followed, is that the child has conceived of the parents as godlike, that is, as the perfect moral agents whose actions should be emulated. We may ourselves remember when we were little children this sense of our parents as almighty beings who could do anything that we wished, or we may remember our own little children who expressed in words to us their expectation that we were almighty beings who could do anything that they wished. This fundamental experience of the weak and helpless child who envisions the parents as almighty in being able to provide for all of the child’s needs is the root of the adult’s concept of God. In *Totem and Taboo*, Freud reinforces this account of the origin of the idea of God when he writes that historically, God arose from a fear of the unknown and incomprehensible, yet the idea of God allowed us to understand or comprehend

An Evaluation of Freud’s Concepts of Morality and God

a surrogate object in God. Thus, the dialectic of original fear established a subject in relation to an unknown, feared object, yet that feared object was eclipsed with the counter-object of God and thus transformed the original subject from fearful, child-like adult into a less fearful, more confident adult through the fantasy of a protective God.¹

Even the adult human being can have profound moments of weakness and insecurity before the powerful forces of nature. But the adult knows now that his own parents are limited and cannot protect their adult offspring from these forces. So the adult in his weakness remembers the false image that the child has of the omnipotent parent and now projects this memory into an equally false image of God in heaven. The two key factors in this exaltation of the child’s memory of the parents into a God in heaven are, first, the continuing weakness even in the adult against the powers of nature and, second, the emotional strength of that original experience of the parents.²

Freud understands religion as having three functions. First, it gives humans a purported explanation about how the universe began, that the divine is the creative source of all. Second, religion offers humans security in an insecure universe and assurance of a final destiny through some reunification with the divine. Thirdly, religion commands humans how they should think, desire, and behave both in relation to the divine and to their fellow humans. Science can be used to serve these same three functions of knowledge of the universe, of protection in an insecure environment, and of moral guidance through an understanding of what behaviors are consistent with promoting a more lasting human happiness. First, science can offer knowledge of how the universe began, for example, through contemporary physics. Second, science can offer some protection against the dangers of nature, for example, through development of vaccinations against contagious diseases. However, science cannot offer the illusion of a perfect reunification with the divine in a heaven. Third, science can offer some moral guidance, based upon what will truly satisfy the longings of human nature, but not with the certitude and definitiveness found in the sacred texts of religions.³

Freud analyzes why religion combines these three functions of knowledge of the universe, of protection and consolation in insecure universe, and of moral guidance of our present behavior. As to the first function of knowledge, Freud argues, that:

... our enquiry is made easy because this God-creator is openly called Father. Psycho-analysis concludes that he really is the father, clothed in the grandeur in which he once appeared to the small child. The religious man’s picture of the creation of the universe is the same as his picture of his own creation.⁴

Once we understand the origin of the idea of the Creator, God the Father, as a projection of the child’s fantasy of the human father as an almighty being, we

can see why the other two functions of religion occur together. For the parents, the father and mother combined as one, are the source of protection of the helpless child in an insecure world and offer assurance of reward and happiness in the future only if the child will behave according to the ethical demands of parents which the child has incorporated in the child’s own super-ego. This whole experience of the child of his origin from the father and mother, his sense of security as rooted in their quasi-omnipotence, and his need to obey their moral demands as the condition of his happiness in the future is transferred over to religious belief almost perfectly, often with a strongly rooted sense of love, Freud notes. First, God is the creative source of all. Just as the human parents have conceived the child in love, so God loves each individual human. Second, God is the source of ultimate security and happiness. Just as the human parents offer security out of love for the child, so also God offers security in this life and the next out of love. Furthermore, just as a child can lovingly petition parents for help, so also adult humans can petition God for help and thereby share in the illusion of participating in God’s omnipotence. Third, God demands the following of moral rules as the condition for security in this life and the next. Just as human parents require moral rules as the guidelines for how to live and love in relation with others, so also God requires moral rules as the guidelines for how to live and love in relation to humans and to God.1

Freud, of course, rejects this whole religious Weltanschauung and affirms that we should adopt the scientific world-view. He views the scientific spirit as challenging the most emotionally powerful elements of the three functions of religion. When humans do follow the moral commands of their deity, religious belief does not give protection to humans from inanimate nature, and it is not true that human virtue is rewarded and vice punished in this life. Rather wickedness and viciousness have held their sway at times throughout history. However, even all the dramatic power of the explanation of the origin of the idea of God by psycho-analysis does not prove, Freud admits, a refutation of religion. Nevertheless, Freud rejects religion as a futile effort to gain power over our frightful perceptual world, in which we live, by means of an illusory wish-world, rooted in our own infantile helplessness and toward which our inevitable projections lead us.2

Karen Horney’s Analysis of the Super-Ego

Karen Horney’s evaluation of Freud offers us a distinction between moral values rooted in rationality and moral values rooted in the compulsive behavior determined by emotional problems in the super-ego. She argues that Freud’s theory was rooted in his analysis of such behaviors as follows:

[C]ertain neurotic types seem to adhere to particularly rigid and high moral standards; the motivating force in their lives is not a wish for

---

happiness but a passionate drive toward rectitude and perfection; they are ruled by a series of “shoulds” and “musts”—they must do a perfect job, be competent in divergent fields, have perfect judgment, be a model husband, a model daughter, a model hostess, and the like.1

Such people, whom Freud describes as ‘obsessive neurotics,’ for example, Little Hans and Ernest Lanzer, have an inexorable obsession towards their goal. Their ideals are so perfect that they must inevitably fail to attain these goals. But because they cannot let go of their compulsive goals, they judge themselves harshly for their present and their past failures. Even if they have been exposed to difficult conditions that affected them as children, they do not see their own responsibility for their failures as lessened. They believe that they should have been strong enough to overcome this maltreatment even as children, but it is their own guilt feeling as children, developed through their identification with the harsh super-ego of their parents, that has trapped them in a cycle of perfect ideals, inevitable failure, and compulsive guilt how in adulthood this sense of helplessness continues and anxiety.2

These perfect ideals fall upon these neurotic types like a curse from the super-ego of the parents, commanding the impossible. Precisely because of this imperative nature of these ideals, Freud himself describes them as ego-alien. He appears to mean, Horney affirms, that these ideals impose themselves upon these neurotics whether these are appealing to them, whether they are justified in a rational analysis, or even whether they can be applied in practically wise ways to the varying circumstances of their lives. Whenever these neurotic types deviate from their impossible standards of perfection, they compulsively go over and over their conduct in their mind in an attempt to rationalize their behavior; but if they cannot rationalize it, their behavior causes them guilt and an anxiety that forces them to reaffirm their impossible goals over and over again, thereby creating an endless cycle.3 We could imagine that in the case of Ernest Lanzer, known as ‘the Ratman’, that the cycle of anxiety/masturbation/guilt/intensified anxiety stems precisely from this ego-alien component Horney has considered, in that the actual voice of the Father seems to command both the prohibitions which cause anxiety, as well as the guilt that rains down when those prohibitions are transgressed through masturbation. Overall, it is this ego-alien ‘voice of the Father’ that is excessively present in Lanzer and which causes him to go through this endless, self-defeating cycle.

Horney concludes that Freud’s main thesis is that the cruel and sadistic way in which the super-ego treats the self is really an inversion of the cruelty and sadism which would be directed against others, but this external aggression is forbidden by the superego and finds its outlet in a torment against the self. Thus, an internal repression is created between the frustrated and repressed id and the highly developed super-ego. Even worse, Horney points out that these super-ego driven people are just as demanding, just as contemptuous, just as ready to condemn others as they are themselves. In support of her point,

References:
2 Karen Horney, New Ways in Psychoanalysis, pp. 207-208.
3Karen Horney, New Ways in Psychoanalysis, pp. 208-209.
Horney notes all the horrors committed against others in the name of religion throughout the ages.¹

In evaluation of Freud’s theory, Horney raises the question of whether the relentless, obsessive, and compulsive need for striving for perfect ideals is really the result of this inner agent, called by Freud the super-ego. Is there not something phony about a demand for perfection which cannot ever be realized and which plunges the self into a spiral of guilt? Is there not something contradictory within the neurotics which, in apparently seeking perfection, nevertheless drives these people into attacks both upon themselves and others? Horney argues that the superego is not the source of striving towards perfection, but rather it is the source of striving for the appearance of perfection. These neurotics only go through the motions of acting with the virtues which they pretend to possess. For, when they fail in their attempt to be virtuous, they do not make a serious attempt to uncover the root of their failure. For example, if they find themselves judgmental towards others even when they know that they should not be, that their behavior is improving neither themselves nor others, they do not make a constructive effort to discover from where this judgmental attitude springs and to work on that problem. They only reassert impossible ideals and continue to be judgmental. But a truly virtuous person will seek the roots of the problem and strive to work on the behavior attempting to develop an Aristotelian Golden Mean in their judgments upon others and themselves that would truly benefit others and the self in developing a reasonable and balanced satisfaction of human needs and well-being. In contrast, these neurotics will excuse their judgmental attitude, arguing that the abysmal failures of others demand these harsh judgments. If this excuse does not work, these neurotics will blame themselves without mercy, rededicate themselves to their impossible ideals, and only create an endless repetition of impossible ideals, inevitable failure, harsh blame of the self, and a rededication to the impossible ideals. When such neurotics come for therapy, Horney points out, they wish that they did not have to come, that they only come because they are suffering phobias or hypochondriac fears or other problems, but they only want their symptoms erased from their lives without their inner personality being transformed.² The conclusion that Horney draws is that such neurotics are driven by a need to appear most perfect, but not by a need for a truly increasing perfection as a state which is better than one’s previous state but not an actual state of perfection. Such neurotics have an insatiable desire to satisfy what is expected of them by others, most precisely that of the harsh and demanding super-ego of the parents.³

Horney concludes:

The whole concept of the “super-ego” is fundamentally changed if we regard the individual’s efforts as directed toward a “pretense” of perfection and infallibility, which for some reason it is necessary to maintain. The “super-ego” is then no longer a special agency within the

³Karen Horney, *New Ways in Psychoanalysis*, pp. 213-216
“ego” but it is a special need of the individual. It is not the advocate of moral perfection, but expresses the neurotic’s need to keep up appearances of perfection.¹

Freud views the super-ego most especially as the source of moral prohibitions and generally as the source of moral ideals. Hence, he believes that the super-ego is fundamentally the same as the normal experience of conscience and ideals, but in an exaggerated form. Even so, for Freud, both super-ego moral demands and normal moral demands are both basically an expression of cruelty against the self. Why, we may ask. It is so because the ego has originally developed an imposing, super-ego ideal, one that harshly prevents the ego from developing harmonious moral ideals integrating self and others into a community of persons both worthwhile in their own sake and also for each other. The obsessive self has never been able to exorcise from consciousness the internalized ideal that is ‘the voice of the Father.’ Thus, the obsessive neurotic’s appearance of being moral would not be for others, so to speak, but for that central other that exists in their consciousness, the ego-ideal of the Father. However, Horney has distinguished more carefully between the ego-alien character of super-ego ideals and the ego-integrated character of one’s moral ideals that would call the present self to transcend itself towards a better self in whom the deepest desires and longings of the present self would be harmoniously actualized with those of other moral agents. The pseudo-moral ideals of the super-ego make it difficult for the real moral ideals of the ego to develop. For example, with the ideal of truth-telling, Horney writes, the super-ego dominated individual superficially observes the standards of truth-telling but experiences this ideal—unconsciously—as an onerous obligation. Only when the neurotic’s truth-telling has ceased to be an alien ideal imposed upon the self from others, only then can truth-telling become a freely chosen ideal that integrates the moral self into the moral community.²

A Response for Freud to the Analysis of Horney

To affirm that the super-ego demands are the same as moral values is Freud’s explicit position, although his own principles more fundamentally require him to deny that position. We have seen Freud assert the equivalence of super-ego and conscience in the material above.³ However, Freud’s principles of mental health affirm:

... that the therapeutic efforts of psychoanalysis ... [are] to strengthen the ego, to make it more independent of the super-ego, to widen its field

¹Karen Horney, New Ways in Psychoanalysis, p. 216.
²Karen Horney, New Ways in Psychoanalysis, pp. 216-217. See also the affirmation by Bruno Bettelheim that it is: “Freud’s conviction that the good life—or at least, the best life available to man, the most enjoyable and most meaningful—consists of being able to find meaningful and satisfactory work that will have positive results also for others.” Bruno Bettelheim, Freud and Man’s Soul (New York: Alfred A. Knopf, 1983) p. 102.
³Sigmund Freud, New Introductory Lectures on Psychoanalysis pp. 88-89.
of vision, and so to extend its organization that it can take over new portions of the id. Where id was [and super-ego], there shall ego be.1

In the last chapter of The Future of an Illusion, Freud develops further this emphasis upon the ego and its control over the id and super-ego. There he proposes that humans should choose the moral ideal of living only in accord with our pursuit of scientific knowledge, only in accord with rational knowledge whose hypotheses can be tested experimentally by a community of investigators all using the empirical method. He names this ideal “our god Logos,” writing:

Our god Logos is not perhaps a very powerful one; he may only fulfil a small part of what his forerunners have promised. If we have to acknowledge this, we shall do so with resignation. We shall not thereby lose interest in the world and in life. . . . We believe that it is possible for scientific work to discover something about the reality of the world through which we increase our power and according to which we can regulate our life.2

The Stoic philosophers had developed the theory of Logos, affirming that there is a Divine Fiery Logos which is source of natural order and laws, called logoi, and which is the source of human reason, human logos, which is a spark of the Divine Reason. Letting go of the concept of Divine Logos, Freud nevertheless affirms the Stoic vision of a natural order governed by natural laws, which human reason can understand. Just as in this Stoic tradition of natural moral law philosophy, human reason can understand the natural equality and dignity of human reason in each person, so also Freud affirms a purely humanistic moral ideal of the dignity of human reason in each person. The task of living according to reason and according to the rational insights of psychoanalysis is to replace control of our individual and social-political lives by the id and super-ego and to develop ego-rational fulfillment in all aspects of our lives.

Consequently, just as Freud himself has argued that the moral ideal should be independent of the super-ego ideal, so also we may argue that the concept of the divine should be independent of the Oedipal complex and the superego. Freud himself completely rejects religious belief in favor of scientific rationality, but his own judgment allows that religious concepts of God fail either to be proven or to be disproven. He much prefers the gradual discernment of the laws of nature being still gradually revealed through scientific hypotheses experimentally verified by a community of investigators. On the one hand, the evidences of our moral lives and our deep wish for God’s all embracing love of our moral efforts and, on the other hand, the evidences of scientific investigations of a purely natural order do not force us to adopt either religious belief or atheistic naturalism. The evidences remain incomplete, but lack of knowledge is lack of knowledge, and Freud concludes that such

1Sigmund Freud, New Introductory Lectures on Psychoanalysis pp. 111-112.
An Evaluation of Freud's Concepts of Morality and God

ignorance does not justify the leap of religious faith.\(^1\) While granting that Freud has developed a powerful argument of living according to the ideal of the god Logos, according to natural human reason in a purely natural order, Freud himself has also offered a powerful critique of basing religious belief in the super-ego-rooted concept of God, permitting the possibility of basing religious belief in an ego-rooted concept of the Divine.

Bibliography


\(^1\)Sigmund Freud, *The Future of an Illusion*, pp. 58-71; and Sigmund Freud, *New Introductory Lectures on Psychoanalysis*, pp. 229-249. The author of this essay expresses his appreciation to Daniel McDow of James Madison University for his analysis of and suggestions for improvement to earlier drafts of this essay.