

Sophia

SOPHIA
UNDERGRADUATE JOURNAL OF PHILOSOPHY
VOLUME XI
2008

Sophia is a non-profit journal, first published in 1998, intended to provide a forum for undergraduate students to exhibit the toils of their undergraduate years. Published annually by students at the University of Victoria, *Sophia* provides a forum for national and international students to present their philosophical thought. It aims at fostering an environment that encourages philosophical discourse and, although its content is principally oriented toward students of philosophy, endeavours to be accessible to all disciplines—philosophy being, after all, not merely an activity for philosophy students alone.

All constructive criticism and comments concerning *Sophia* or any of the published works therein are welcome and encouraged. Information for students interested in submitting work for future volumes can be found on the last page of this journal.

CHIEF EDITOR
EDITORS

GEORDIE MCCOMB
PAMELA ROBINSON
LINDSAY BOURQUE

GRADUATE READER
COVER DESIGN

MARY BUTTERFIELD
ALLISON PARKER

This year's edition of *Sophia* had an unusually high number of submissions, the vast majority of which came from outside the University of Victoria. For the completion of this year's edition, we are grateful to many people not listed in the above credits: all those who submitted to *Sophia*, the fundraising efforts of the members of the Philosophy Students' Union, the support from many faculty members, funding from the University of Victoria Students' Society and the Canadian Philosophical Association for posting our call for papers on their website.

Copyright of published work remains the sole property of the authors.

TABLE OF CONTENTS

An Interview with Allen W. Wood	1
Formulating Maxims: Defending the Formula of Universal Law Against False Negatives and Positives CHRIS PERKINS	15
Special Relativity and the Time-Lag Argument JOHN KARDOSH	23
Judging the Judges: Wittgenstein's Sceptical Paradox for Debates in the Philosophy of Neuroscience WILL ROBBINS	33
Foucault and Freedom: On the Constitutional Efficacy of Practice ADAM MICHIELS	47
A Metaphorical, Heuristic Account of Aristotle's Doctrine of the Mean BENJAMIN VISSCHER HOLE IV	59
On Late-Capitalist Desire BRENT VIZEAU	71
The Quest for the Virtuous Life: Can Lao Tzu's Wu Wei be Considered a Virtue? CHRIS DIMATTEO	81
Relevant Possibilities, Intentions, and Moral Responsibility DANIEL MOOSE	89
Defining Goodness SIMONA AIMAR	101

An Interview with Allen W. Wood

Allen W. Wood is Ward W. and Priscilla B. Woods Professor in the Department of Philosophy at Stanford University. His research and interests focus on the History of Modern Philosophy (especially German philosophy in the eighteenth and nineteenth centuries); Ethics; Social and Political Philosophy; and Philosophy of Religion. His books include *Kant's Moral Religion* (1970), *Kant's Rational Theology* (1978), *Hegel's Ethical Thought* (1990) and *Kant's Ethical Thought* (1999).

During his stay in Victoria this past September as a presenter for the University of Victoria's Landsdown Lectures in the Faculty of Humanities, we were honoured that Dr. Wood agreed to be interviewed on his most recent book, *Kantian Ethics* (2008)¹, for this edition of *Sophia*.

I would like to thank Lindsay Bourque and Pamela Robinson as well as Dr. Colin Macleod for their helpful suggestions.

Geordie McComb

¹ All page references in the following text are from this book: Wood, Allen W. *Kantian Ethics*. Cambridge, UK ; New York : Cambridge University Press, 2008.

Sophia: In the preface to *Kantian Ethics* you write that you partially fulfilled the late Terence Moore's request to provide a briefer, less scholarly and more approachable version of Kantian ethics. We might think that a book even partially fulfilling such criteria would draw the interest of a broader readership. If this is right, why is such a book needed and what would you hope for it to achieve?

Allen Wood: I can't comment on what the readership of my book will be or should be. That's for the potential readers to decide, not authors (or publishers). But the book is intended for people who already have an interest in philosophical reflection on ethical issues and at least a curiosity about what a Kantian approach to these issues might be. What I hope to achieve, in the case of anyone who may choose to open the book and read it, is a change in the common image of Kantian ethics that they may have picked up from other philosophers, or even from reading Kant's writings themselves. I sketch the way Kantian theory can (and should) be applied to some important moral questions—about personhood, for instance, or the moral status of animals, or the role of love and sympathy in morality, or the justification of punishment, economic distribution, sexual morality, lying, or the relation of principles and consequences in making moral decisions. I try to shed light from a Kantian point of view on the proper aims and structure of a moral theory and the way moral theories relate to ordinary moral decisions. I also have things to say about Kantian views about autonomy and freedom that go against some prevailing notions. I think some fairly radical revisions are needed in the ways people think about Kantian ethics if there is to be an accurate and philosophically useful picture of Kant as a moral philosopher.

S: In *Kantian Ethics* you frequently point out places where many interpreters of Kant's ethics go wrong. Many interpretations, for instance, place too much weight on Kant's *Groundwork for the Metaphysics of Morals* and this work is often taught in undergraduate philosophy classes at the expense of Kant's other major ethical works, such as his *Metaphysics of Morals* and his *Critique of Practical Reason*. Yet, given the time constraints on teaching Kant's ethics in such classes, the length of the *Groundwork* seems appropriately short. Given such issues as time constraints and common misinterpretations of Kant's ethics, how should undergraduate students approach learning Kant's ethics?

AW: When I teach Kant's ethics, I always begin with the *Groundwork*,

and it is the only Kantian text I teach in courses on ethical theory, where there is time only for a brief introduction to Kant. So I am not the least bit opposed to the idea that people who want to study Kant's ethics should read the *Groundwork* first. Clearly Kant intended that they should.

What I want is only that people should understand that work better. I try to read the *Groundwork* in ways that exhibit awareness of its limited aims and that do not distort Kant's views about other things. The *Groundwork* (as well as the *Critique of Practical Reason*) is intended solely as a foundational work. You should not try to get the Kantian answers from it to questions that are better answered by some of Kant's other writings. People also often come to the *Groundwork* with false assumptions drawn from ideas about ethical theory that are more fashionable today, but not shared by Kant. This leads them to take away from the *Groundwork* a set of mistaken ideas about (for instance) the fundamental values on which morality rests, the structure of moral theory, the distinction in Kant between right and morality, and the ways ordinary moral agents should make decisions. Other common errors are understandable in light of what the *Groundwork* does not say on certain subjects, such as the foundations of right or certain questions of moral psychology that are dealt with in Kant's later works. But I think you won't understand Kant, or even the ethical issues, as well as you should until these errors are corrected.

S: One such common misinterpretation you consider is an understanding of the Kantian moral agent's relationship with other people as coldly indifferent, detached, unconcerned and emotionally repressed. In your discussion of the Kantian classification of duties, you present Kant's distinction within our ethical duties to others: duties of respect and duties of love. One particularly striking duty of love is humanity or "sympathetic participation"—an emotional duty insofar as we should cultivate certain feelings for people but also a cognitive duty insofar as we should try to understand, be involved with or be empathetic towards these same people. One difficulty we might have with understanding this classification of duties is that properly speaking love, for Kant, is "pleasure in the perfection of another leading to a desire to benefit the other" (179). How is it that sympathetic participation is a duty of love when it appears to have little to do with a part of what "love" means—namely, taking pleasure in the perfection of another?

AW: It is not true that sympathetic participation has little to do with love

in the sense of a taking pleasure in the perfection of another that leads to a desire to benefit the other. Kant gives the name “practical love” to the maxim or policy of benefiting another, and distinguishes this from “pathological love” (by which he does not mean there is anything sick about it, but only that it is love based on empirical feeling—Greek *pathos*). He holds that when we show practical love to another, this tends to produce in us love as feeling—love as pleasure in the perfection of the other leading us to benefit the other (*Metaphysics of Morals* 6:402). Sympathetic participation, taking a caring or empathetic stance toward the other, would in Kant’s view be part of any practical love that was genuine and not just hypocritical show. So sympathetic participation would be part of the policy that leads to love as feeling. Conversely, love as feeling, if it is genuine, would also give rise to sympathetic participation. There is also a third kind of thing Kant calls “love”—namely, “philanthropy” or “love of human beings.” This is a non-empirical feeling (a feeling arising directly from the influence of reason on our sensibility). Along with such feelings as moral approval, conscience and respect, Kant thinks love of human beings is a basic part of our makeup as rational and moral beings (*Metaphysics of Morals* 6:399-403). Without these feelings, we could not have moral obligations or be complete as rational agents at all. To fulfill the duty of sympathetic participation would normally be to act the feeling of love of human beings, and that action, once again, would normally produce in us love as empirical feeling for the person. So all the things that Kant calls “love,” though they may be distinct in their definitions, are closely connected in their psychology.

S: At several points in *Kantian Ethics* you dismiss an understanding of rationality that reduces it to “technical” or “instrumental” reason because it leads to nonsense or patent absurdities: “Few philosophers have been as candid about this as Hume was when he boldly asserted such patent absurdities about reason as that it can never motivate the will and that it is not contrary to reason to prefer the destruction of the world to the scratching of my finger” (15). Specifically, this view of rationality, what is often called instrumental rationality, has reason determine what we ought to do given certain ends but never what ends we ought to pursue. Why is it that instrumental rationality—if, as you write, it is absurd—seems to enjoy acceptance among so many philosophers?

AW: Usually in any widespread philosophical error, there is an element of truth—and the error consists in having misunderstood this truth, perhaps exaggerating its importance, allowing it to obscure other truths,

or misinterpreting its importance in relation to them. For Kant, the most basic act of practical reason is setting an end, and the concept of an end contains in it the concept of taking some means to the end. (A desire that never rises to the thought of making its object an end and taking means toward it is merely an idle wish.) So instrumental reason figures somehow in every action based on practical reason. But Kant thinks there are other aspects of practical reason too. There is prudential or pragmatic reason, concerned with choosing which ends to set so that they form a coherent whole, constituting the welfare or happiness of the person (normally, yourself) for whose sake the ends are being set. And then of course there is also moral reason, concerned both with which ends you ought to set irrespective of your own welfare, and with principles that ought to be followed prior to the setting of any ends, which even determine something about the ends you should set. The necessary moral ends, in Kant's view, are your own perfection and the happiness of others. And principles binding on you prior to and irrespective of any ends are those involving respect for the dignity and the rights of persons.

Without instrumental reason, there could be no use of practical reason at all, but that doesn't mean there is nothing to practical reason except instrumental reason. Analogously, satisfying the basic conditions for physical survival is the necessary condition for having any life at all, but there is a lot more to life—and especially to a meaningful and flourishing life—than bare physical survival. Someone who thought that physical survival is the only end of life would have an impoverished view of what life is about, and in the same way, someone who thinks there is nothing to practical reason except instrumental reason has an impoverished idea of what practical reason is.

Why do people come up with such impoverished conceptions? One explanation may be that it seems controversial what principles should govern them in ways that they think it is not controversial how principles of instrumental rationality work. Standard theories of practical rationality find quite enough controversy right in instrumental rationality itself, though they think these controversies can be resolved satisfactorily.

A second explanation is that they may be under the illusion that everything we need or want that pertains to prudence (or to morality, insofar as it may be reasoned about) can be brought under the principles of instrumental reason—usually, plus a few “accounting” principles having to do with consistency in the setting of ends and setting coherent priorities among them and the use of means to them. Prudence and

morality, when morality is regarded as the source of reasons and rational principles, obviously open up larger philosophical controversies, that some philosophers would prefer to avoid. Not only do they raise questions about fundamental moral disagreement, but prudential and moral rationality have uncomfortable relationships to other vexed philosophical questions, such as freedom of the will or the metaphysics of value (now commonly called ‘metaethics’). I think many philosophers would like to think these latter questions can be dealt with in ways that are non-metaphysical or metaphysically uncontroversial—for instance, by adopting standard compatibilism about free will, or noncognitivist and anti-realist views in metaethics.

Large claims about moral reason or the human good, when they can’t be reduced to tidy theories of rationality focusing on instrumental reason, may seem impossible to unite with “naturalistic” views about ourselves and the world. The restriction of reason to instrumental reason therefore looks philosophically tidy and modest, and it avoids the risk of having to deal with theories of volition and value that seem too hard to reconcile with other philosophical commitments, or of having to admit we don’t understand these matters as well as we would like to pretend we understand them. As I say in the book, I don’t understand Kantian ethics as committed to supernaturalistic theories of the will, for instance, but I do think it points to ways in which many standard empiricist and so-called ‘naturalist’ theories of these matters are unsatisfactory. Kant taught that human reason is beset with many questions it cannot help raising but to which it must admit it can provide no satisfactory answer. The version of Kantian ethics I favor also involves this admission about questions of free will and the relation of values to what is real. (I discuss these matters in Chapters 6, 7 and 15). I don’t see Kantian ethics as committed to a supernaturalist metaphysics, but I do see it as raising problems for some standard naturalistic views, and as confronting us with the fact that the human condition may in some ways be incomprehensible to us in ways that both naturalists and supernaturalists do not want to admit.

As a response to the perplexity and absurdity of our condition, I see both naturalism and supernaturalism as exhibiting states of dishonest denial. I think naturalistic theories are the only ones we should seriously listen to, but we should be ready to admit when they give out and not try to use them to answer insoluble philosophical questions. This seems to me the fundamental Kantian message about such matters, though I am aware that there are alternative interpretations of Kant, with some textual support, that take a line more hospitable to supernaturalism. I regard the

strands in Kantian texts that support these alternatives as inconsistent both with the central message of Kant and with plain human honesty and good sense.

S: Regarding this admission in your understanding of Kantian ethics, that the question of free will cannot be satisfactorily answered by human reason, you write that the right way to think about Kant's metaphysically extravagant and unacceptable "answer" to this question, his theory of noumenal freedom, is as a painful reminder of the existence of the problem of free will. Indeed, this question is one of those that "endlessly torments us and won't let go, because we cannot, while retaining our self-respect as rational beings, ever let go of them" (141). One concern someone might have with this position is that we typically take an insoluble problem such as this one to be a good reason to revise the thinking that led to this problem. Certain steps that led to the purportedly insoluble problems associated with Cartesian or mind-body dualism, for instance, are often criticized and even rejected on such grounds. Should we not, likewise, count the insolubility of the problem of free will as a serious criticism of Kantian ethics and consider rejecting at least an important part of it on these grounds?

AW: I do not think the problem of free will depends the least bit on accepting Cartesian mind-body dualism, which almost nobody accepts any longer. If it were that easy to get rid of the problem of free will, then we would no longer be thinking about it at all. Kant, by present day standards, was a rather extreme skeptic about the mind-body problem, and thought all the traditional theories about it had insuperable difficulties with them. He certainly did not think the question of free will arises only for those who accept a Cartesian theory on the relation of mind and body.

I think you are living in a fool's paradise if you think that all philosophical problems admit of solution (or "dissolution") simply by locating and excising certain errors on which they are based. That was a fashionable logical positivist view early in the twentieth century, and it was taken up by many philosophers in the analytical tradition, including Wittgenstein. It rationalizes the complacent idea that we can sweep philosophical questions out of the way, and deal only with natural scientific questions, which are all answerable. But the sciences have grown up with philosophy, alongside it and intertwined with it. This means that scientific theories never escape philosophical presuppositions. It also means that scientific knowledge, as well as what we like to call

“common sense,” is forever surrounded with doubts and perplexities we can’t get rid of. We just have to learn to live with them somehow, though they will always make the lives of honest inquirers uncomfortable. Kant put it very well in the very opening sentence of the *Critique of Pure Reason* when he said that certain questions are posed for us by the nature of human reason, which we therefore cannot dismiss, even though they surpass the capacity of our reason to answer them.

If all philosophers face an unsolved, possibly insoluble, problem about free will, then that is no more a defect in Kantian ethics than it is a defect in any other position. Kantian ethics did not create the problem of free will. No rival theory has a satisfactory solution to it. The idea of making choices based on reasons is fundamental to any ethical theory, so no theory can really avoid the free will problem either. If a philosopher faces up honestly to a problem, admits its existence and shows its unavoidability, that is a virtue in the philosopher. It is not a reason for rejecting his other views. Kant seems to me to have this virtue regarding the problem of free will. To twist the virtue into a reason to criticize the parts of his philosophy that relate to the problem seems rather like the practice of certain childish despots who, when a messenger brings them unwelcome news, put the messenger to death. (The Bush administration has given us a kinder, gentler version of this barbaric practice: Bush and Cheney surround themselves with people who agree with them, and dismiss as disloyal anyone who tries to question their decisions or to correct the errors on which those decisions were based. In the past seven years we have seen the disastrous results of that policy.) Just as killing the messenger does not make the bad news go away, so rejecting the philosophy of those who tell you the truth about unsolved problems does not make the problems go away. Those who face up to the unsolved problems are more likely to give you defensible views than those who are in denial about them.

If someone thinks they have a solution to the problem of free will, then they should tell us what the solution is, and when we have heard them out, and are satisfied that they have solved the problem, that will be the time to consider whether the solution favors one particular ethical view over another. I don’t think a satisfactory solution to the problem, if one could be found, would leave Kantian ethics any worse off than it is. In my book, I describe what I think might count as a solution to the problem of free will—namely, a form of compatibilism that shows us how we have genuine options, and a causal capacity to act of the kind that Locke calls an “indifferent power”—a causal power which can be exercised by

choosing either of two ways, so that the natural causal series does not necessitate what we do, but rather provides us with the capacity to act from what Leibniz calls reasons that “incline without necessitating.” I don’t think anyone has provided that sort of solution as yet, but if they did, I think the result would be a naturalistic account (no noumenal selves or other supernatural agencies) of exactly the practical freedom—the ability to act on reasons and from rational principles—on which Kantian ethics is based. I think the problem of free will, and any workable solution to it, is essentially neutral among competing ethical theories. To prove me wrong about that, you would first need a satisfactory solution to the problem—something I do not think anyone has yet provided.

S: In your chapter on ethical theory you raise doubts about whether our moral intuitions are credible data in the moral epistemology of the “standard” model of ethical theory. One way you raise such doubt involves thought experiments such as the “trolley problems.” A trolley problem seems to be a thought experiment in which one imagines oneself ready to flip a switch that changes the lane of an otherwise out of control trolley car. It is further imagined that this trolley car will kill certain people if you flip the switch and others if you do not. One is then asked to say whether or not one should flip the switch—one’s answer being the moral intuition that the standard model of ethical theory takes to be credible data. You object that certain factors that presumably would be the case in the real world are abstracted away from many such thought experiments rendering them incapable of yielding, I presume, moral intuitions *qua* credible data. You cite as important elements that are abstracted away from the trolley car problem, for example, the legality of flipping the switch, the existence of standard operating procedures, the uncertain epistemic position of the switch flipper and being prosecuted for flipping the switch. Now, it might be objected that such elements are abstracted away precisely because they are not morally relevant when we consider such trolley problems. How might you reply to a critic that rejected your doubt about the reliability of thought experiments on such grounds?

AW: When I pointed out that it would be illegal for a mere bystander to flip the switch, I of course meant to suggest that such laws are entirely correct and appropriate. It is not a case of an unjust or wrongheadedly restrictive law, such as the laws that required returning runaway slaves to their owners or “anti-sodomy” laws prohibiting certain sexual acts between consenting adults to which there is no rational moral objection.

We do have overwhelmingly good grounds for prohibiting this by law. There would be no end to the disasters that could be caused by allowing incompetent or possibly malicious strangers to change the tracks on which trolley cars are running. Nobody who thinks about the issue for more than two seconds could possibly think it would be just dandy to leave such matters unregulated. But then how could it possibly be morally irrelevant to what you should do that flipping the switch would be correctly prohibited by law? There might be cases, of course, in which it could be morally justified to break such a law, perhaps, to save many lives. But as trolley problems are usually stated, the issue of the likely illegality of what the bystander is doing is not even raised. Would people's "intuitions" about what a bystander should do be the same if it were an explicit part of the problem that they were likely to be sued or prosecuted if they killed one person to save five? I think in real life, if anyone died as the result of a bystander exercising his moral intuitions by flipping the switch, there would be a big public outcry over the fact that a mere bystander had been permitted near the switch and someone had died as a result. That five lives had been saved might complicate the discussion, but it would not be seen as removing the scandal.

I mention other obvious facts from which the problems usually abstract, such as that on the track where you see only one person standing, there might be a dozen others just out of sight—and that a casual bystanders could never be in a position to know that this is not so. When such inconvenient general facts interfere with the operation of their "intuition pumps," philosophers often react by simply "stipulating" that we know there is no risk to others that we might not be aware of. But you are not creating a realistic example, or a fair appeal to intuitions, if you stipulate that we know things we never could know in real life.

The impatient spirit in which philosophers make such stipulations when you start raising these troubling complications shows that what they are really doing is dealing with rather abstract problems, in which we are supposed to ignore the complicating facts, since the intuitions the philosophers mean to elicit are about abstract questions, not about any actual situation people might face. There are strong theoretical assumptions, however, behind the thought that our intuitions about *these* abstractions are the important ones. Roughly speaking, it looks as if trolley problems assume that at bottom, all moral questions must turn solely on (1) what the specific foreseeable consequences of an act are for human interests and (2) how the act causes those consequences. (As I point out in the book, the importance of (2) in the intuitions elicited by

trolley problems shows that they do not assume a consequentialist stance, and are often used to elicit anti-consequentialist intuitions.)

If my speculation here is right, then the unstated background assumption of trolley problems needs only to be made explicit in order to appear highly questionable. For surely it is not the least bit self-evident that no more is relevant to a great many moral decisions than the immediate foreseeable consequences of an action and how the action produces them. There are also, for example, the consequences of larger scale social policies and practices (under which the action might fall), and also the way the action may conform to or violate norms or values not contained directly in their immediate foreseeable consequences. No theory—Kantian, utilitarian, or any other—can afford to ignore these other factors and to recognize their independence of considerations of the immediate consequences of an action and the manner in which those are caused.

No doubt someone who has (perhaps thoughtlessly or without even realizing it) assumed that only the immediate consequences of actions and the manner of causing them could be morally relevant to the rightness of the action will regard the introduction of other factors as “irrelevant.” But the judgment that they are irrelevant is by no means unquestionable or theoretically innocent. When a trolley problem elicits our “intuitions” about the abstractly conceived situation, it does not give us the opportunity to reject the theoretical assumptions that have been built into the problem. If the question were posed in a way that brought those assumptions to our attention, then we might turn out to have strong “intuitions” that run contrary to them.

It follows, I think, that we cannot take people’s responses to trolley problems at face value as credible data for moral theorizing. In the book I compare this to cases where the results of public opinion polls can be seen to vary depending on the way questions are put in them. For instance, in eliciting opinions about the death penalty, do you mention or not mention alternative penalties for murder, such as life in prison without parole? Do you or do you not provide data bearing on the question how often innocent people are executed, or point out that the relative cost to the judicial system of imposing the death penalty is usually greater than that of incarcerating the murderer for life? On difficult questions, many people have no settled position, and what they say to a pollster may be determined more by what they think the pollster will think of them in giving their answer than on what they think about the issue. They don’t want to appear too opinionated or to be ignoring

things they think the pollster regards as important, since that will make them look hasty or thoughtless in the pollster's eyes. So what they say may depend on what aspects of the issue are mentioned in the pollster's question and what aspects are left out.

Likewise, if all you are given in a trolley problem is how many people will die if you do this or that, and whether in doing this you are actively killing them or merely letting their death happen without actively preventing it, then those are likely to be the only factors your response to which is elicited in your "intuition" about the problem. There may be many other moral beliefs, or potential "intuitions," that would show themselves in your response if the moral issue were framed in ways that brought them to your attention. As soon as you notice that trolley problems are abstracting from factors that would be at work in real life, that should be an immediate tip off that our responses to trolley problems should not be considered credible data for moral theory.

S: At the end of your chapter on consequences, you hold that we have justified ethical convictions and that Kantian ethics, as you understand it, is "the best kind of theory human beings have devised as a background for their moral reflections" (272). Furthermore, you write that such a theory may aid us in our moral thinking by helping us to criticize our ethical convictions and, at times, by leading us to form new ones. We might think that such a situation should lead us to a belief in Kantian ethics. You argue, however, that Kantian ethics as an ethical theory is deeply questionable and controversial. We will never, moreover, achieve certainty about the rational basis of our ethical convictions nor are we even justified in believing any philosophical theory about ethics, including Kantian ethics. Given such a view of Kantian ethics as uncertain, questionable and not apt for justified belief, how can we legitimately criticize our partly justified ethical convictions?

AW: I think that all ethical theories are uncertain, questionable, and not apt for justified belief. Many of them, however, are useful in thinking about the ethical convictions we do believe in, and in changing those convictions. If used for that purpose, they don't have to be considered "true" or "believed in." All we need to do is adopt them for certain purposes as working assumptions to see whether they provide insight on the questions we are trying to think about. I do *believe* that Kantian ethics is the most useful ethical theory for this purpose. But that does not entail *believing Kantian ethical theory*, any more than it gives me a reason to *believe* in competing ethical theories, though I may often find

them useful too, for thinking about moral problems, or even thinking about the relative strengths and weaknesses of Kantian ethics.

Your question also seems to presuppose that in order to use a set of thoughts to justify, criticize and change our beliefs we have to believe in the thoughts. I think the experience of doing philosophy shows that this is not true. Philosophers often try out arguments, thought experiments and theoretical constructions whose claims and premises they would be foolish to *believe*, but they reach conclusions from them—sometimes conclusions they even come to *believe*. It seems to me that if we always had to start out in philosophy from a set of beliefs that are well justified, then we could never begin to deal with basic philosophical questions, since these are always perplexities which stand in the way of having justified beliefs about the foundations of morality, or knowledge, or theories about what there is. Or, if we were always in a position to address such questions from the standpoint of well justified beliefs, then I don't think they would be what we now call *philosophical* questions. In fact, I don't think there would be any properly *philosophical* questions at all.

Here I come back to my earlier point that philosophy reflects the human condition, which is one of irresolvable perplexities and unanswerable questions. To be a philosopher is to be deeply uncomfortable with who you are, what you are and where you are (in this vast, incomprehensible and ultimately meaningless universe). But then every human being who is honest is a philosopher in this sense. To use science, or religion, or even just plain thoughtlessness, to avoid the discomforts resulting from the irretrievable absurdity and unconsolability of the human condition is to be guilty of a lack of intellectual integrity. It is both contemptible and blameworthy. The existentialists understood very well that flight from the absurdity of the human condition is among the most unforgivable things a human being can be guilty of, and it is also the cause of many other unforgivable crimes—such as those resulting from some project of forcing other people to see the world the way you see it, in the desperate attempt to confirm in your own mind some faith or ideology that helps you escape from your own fundamental and unavoidable discomfort as a thinking human being.

S: You have relatively recently published two books on Kant's ethics, *Kant's Ethical Thought* and, as already mentioned, *Kantian Ethics*. Where do you see your work going from here?

AW: For many years now, I have been telling myself (and sometimes other people too) that, having written a book on *Hegel's Ethical Thought* and a book on *Kant's Ethical Thought*, I should next write a book about the important figure that came between them historically—that is, a book on *Fichte's Ethical Thought*. But since I first said that, I have written four other books, so maybe I should stop predicting I will write a book on Fichte and just admit frankly that I am not very good at predicting where my work will go from here. Maybe if I do that (who knows?) I might someday end up writing *Fichte's Ethical Thought* after all.

Let me end by thanking you for your interest in my work and for giving me the opportunity to talk a bit more about Kantian ethics.

Formulating Maxims: Defending the Formula of Universal Law Against False Negatives and Positives

CHRIS PERKINS

UNIVERSITY OF VICTORIA

In this paper, I discuss some counterintuitive results of Kant's Formula of Universal Law (FUL), known as false positives and negatives. These false results are a critical problem for FUL if they cannot be avoided. I argue that the cause of at least some of these false results is the improper formation of maxims. Rules regulating the creation of maxims would help to prevent false results, and I attempt to develop one such rule. To do this I investigate examples of false results drawn from the work of Allen W. Wood and identify flaws in their maxims. I arrive at a rule that specifics like names, dates and times should not be included in maxims. This is used to show that at least one justifiable rule exists that can eliminate some false results.

Kant's Formula of Universal Law (FUL) has been interpreted as a rational algorithm that can determine the moral worth of actions. FUL depends on the concept that people act on *maxims*—practical rules that dictate the best way to accomplish given goals. FUL (and the test derived from it) judges actions based on the nature of the maxims that determine them. Kant does not provide explicit instructions for extracting maxims from actions, nor does he say exactly what a maxim should look like. This ambiguity creates serious interpretive problems.

Critics of FUL have pointed out cases where the test assigns an action a moral status that is contrary to popular opinion. These cases are known as “false negatives” and “false positives,” and they are fatal to FUL if they cannot be prevented. At least some of these false results are caused by improper formulation of maxims. There is both a need and justification for criteria governing the formation of maxims. In this essay I will give an explanation and justification of one criterion that I have devised: that specifics like names, dates, places and times, should be left out of maxims and replaced by more general conditions. My hope is that this attempt will validate a search for other maxim-governing rules which collectively prevent all false results, thereby absolving FUL of a damning charge.

1. The Formula of Universal Law and the Universalization Test

In Immanuel Kant's *Groundwork for the Metaphysics of Morals*, the Formula of Universal Law is stated thus: “Act only in accordance with that maxim which you can at the same time will that it become a universal law” (37). This formula depends on Kant's concept that rational beings decide how to act based on practical rules. These rules, called “maxims,” tell us how to accomplish our desired ends, and we choose maxims based on how effectively they achieve those ends (16).

If we know what maxims guide our actions, we can judge our actions' moral status using FUL. The formula dictates that we imagine a world in which everyone made decisions based on the same maxims that we do, and ask ourselves if we can will that world to exist. Kant claims that when we universalize an immoral maxim, some sort of rational contradiction will be created, so that we cannot possibly will that maxim to be made universal law (38-40).

Some Kantians believe that the moral status of actions can be determined by a systematic test, derived from FUL, which I call the Universalization Test (UT):

1. Determine the maxim of the action in question
2. Imagine a world where everyone follows that maxim
3. Look for contradictions within that world

4. If a contradiction exists, then the action fails UT and is morally impermissible
If one does not exist, the action passes UT and is morally permissible

Kant does not explain what exactly constitutes a contradiction, so I will adopt Christine Korsgaard's account of "practical contradictions" from her book *Kant's Moral Philosophy*. In this interpretation, an immoral maxim is "self-defeating if universalized," which means that the action in question would not achieve its purpose if everybody did it (78). This interpretation is certainly adequate for the purposes of this essay.

2. False positives and negatives are potentially fatal for UT

False positives and negatives are generally described as results of UT that assign a moral value opposite to popular belief. Kant scholar Allen W. Wood believes they prove FUL to be an "incomplete or defective representation of the fundamental principle of morality." In my opinion it is not problematic if UT contradicts convention (see Appendix). It is more problematic that false results contradict other, more acceptable results of UT. As it stands now, UT allows for multiple maxims to be derived from a given action. This is a serious problem when some maxims for an action pass UT while others fail. For example, Allen Wood and Kant both discuss a case of fraudulent borrowing. From this action Wood derives a maxim that passes UT, implying that fraudulent borrowing is permissible (102). Kant derives a different maxim that *fails* UT, and concludes that fraudulent borrowing is *impermissible* (39). I believe both writers performed UT correctly on the maxims they chose. If both maxims are acceptable, then UT produces contradictory results. Presumably an action cannot be both permissible and impermissible, so this constitutes a *reductio ad absurdum*. Therefore either UT is inherently flawed, or one of those maxims is not acceptable.

3. Creating criteria for maxims is crucial for defending UT

The problem of false positives and negatives is fatal for UT unless some maxims can be deemed unacceptable and rejected. Without criteria for evaluating maxims, we cannot make claims about acceptability. Therefore it is necessary for those who wish to preserve UT to develop criteria governing the formulation of maxims. I believe that UT is worth defending, so I will attempt in this paper to develop one useful criterion for maxims. Hopefully this will validate a search for a more complete set of rules. This pursuit is particularly worthwhile because it will help save UT from internal contradiction, but in many cases the contradictory results are also counterintuitive and unfeasible.

Therefore, creating criteria that eliminate these results will also make UT more intuitively palatable.

4. Excluding specifics from maxims

In his discussion of false positives, Allen W. Wood discusses a situation where a tennis player wishes to avoid busy tennis courts. The player knows that on Sunday mornings all of his neighbours are in church and thus nobody is playing tennis. He does not attend church, so he decides to play on Sunday mornings when the courts are free. Wood thinks that the decision to play on Sundays is determined by the following maxim:

T1: “In order to avoid crowded tennis courts, I will play on Sunday mornings (when my neighbours are in church and the courts are free)” (105).

If this maxim were made universally compelling, then everybody would play tennis on Sundays and the player’s purpose would be defeated. In a world where everyone played tennis on Sunday to avoid crowds, playing tennis on Sunday would not be a good way to avoid crowds. When universalized, T1 contradicts itself and is therefore impermissible by UT. Because playing tennis on Sunday is an obviously acceptable practice, this result is a false negative.

However, if we word the maxim more carefully, we can avoid this false result:

T2: In order to avoid crowded tennis courts, I will play at a time when I know the courts are least busy.

In the given situation, this maxim recommends the same action (to play on Sunday) but it is *not* self-defeating. Because it specifies a condition rather than the details of day and time, the problem of overcrowding is avoided. If Sunday becomes too busy, then the maxim dictates that we play on another day that is less busy. If everyone who wanted to avoid crowded courts did this, then use of the courts would even out so that nobody would have to play at particularly busy times. T2 passes UT and thus does not create a false negative.

T2 is a superior maxim to T1 because it better expresses the nature of the choice to play on Sunday. The condition “a time when I know the courts are least busy” is more relevant to the choice to act than the detail “Sunday morning” because the detail depends on the condition. Assuming that he is rational, the crowd-avoiding player that Wood speaks of only chooses Sunday to play because that is when his neighbours are in church and the courts are free. If the courts were full on Sundays and empty on Tuesday mornings, playing on Tuesday mornings would be a much better way to avoid crowds. Trying to avoid

crowds by playing on Sundays, no matter what, would be an impractical policy. This is exactly what T1 is recommending: “in order to avoid crowds, I will play tennis on Sunday.” T1 only seems like an acceptable maxim because the reasoning behind it is provided. T2 is more appropriate because that reasoning is laid out within the maxim itself.

Kant claims that rational actions are determined by maxims, and maxims are wholly aimed at accomplishing some end. A maxim would not recommend a purposeless action, so every element of a rational action must also be aimed at accomplishing an end. When we are deciding rationally, we choose the details of an action, like names, dates, times and places, because they are useful for accomplishing something. For example, our tennis player chooses Sunday to play because that is when the courts are least busy; thus, he can best avoid crowds on that day. When we are extracting maxims from actions, it should be possible to determine the reason for every detail. If this is the case, then there is no need to include specifics like “Sunday morning” in maxims. We should be able to restate those details as conditions like “a time when I know the courts are least busy” that clearly relate to the maxim’s aim.

Since it is possible to avoid them, maxims should not include the details I have discussed above. Maxims are *general* principles that guide decision-making, and principles that include specifics are not general. “To avoid crowds, play tennis on Sunday mornings” is only helpful advice in the case that courts are least busy on Sunday mornings. T1 fails UT because it specifies a particular time and cannot recommend a different course of action when the courts become busy on Sunday. A general principle should be able to determine the best course of action in a wide range of circumstances, and a principle that includes details is only useful in specific cases. Therefore a principle that includes specific details is not a general principle and cannot be a maxim. From this we can derive a rule:

The Omission of Specifics (OS): When deriving a maxim from an action, specifics of the action such as name, date, time and place should be omitted. If any specifics are relevant to the maxim’s end, the conditions that recommended those details should be included.

To demonstrate the need for, and effectiveness of, this rule, I will apply it to another of Wood’s examples that suffers from the same problem as T1. Wood claims that UT will permit “any maxim that tells us to make a false promise but sufficiently restricts the conditions under which we do so” (102). The maxim he uses as an example is roughly as follows:

F1: In order to acquire money, I will borrow from Hildreth Milton Flitcraft at 3:30 p.m. on August 21, knowing that I will not repay the debt.

Wood believes that there is no contradiction created when we universalize this maxim because it is too specific to make an impact even if everyone followed it. Thus it passes UT and fraudulent borrowing is deemed permissible, which is a false positive.

Wood is not justified in this conclusion. If every person who desired money tried to get some by making false promises to Hildreth at exactly the same time, it would be impossible to secure a loan for the simple reason that her house would be overcrowded. Universalizing F1 creates a practical contradiction, so it fails UT. This deems fraudulent borrowing impermissible and there is therefore no false positive created.

It does not seem appropriate, however, that fraudulent borrowing should be deemed impermissible on the grounds that the world's population cannot fit in one person's house. F1 fails UT for the same reason that T1 does: it is impractical for everyone to do the same thing in the same place, at the same time, with the same person, when resources are limited. Potentially, every maxim involving the use of limited resources will fail UT for this reason, if it includes specifics like names, times and places. Not every maxim that involves the use of limited resources is intuitively wrong, and others are wrong for different reasons. Including specifics in maxims causes purely logistical problems that do not give the maxims a chance to pass or fail on their own merits.

This problem can be avoided by following the Omission of Specifics rule. In the case of fraudulent borrowing, we can assume that all of the details in F1 are somehow relevant to gaining money. Presumably Hildreth Milton Flitcraft is a person with the will and means to lend money. Perhaps the time and date are when she is earliest available. We could restate the maxim as:

F2: In order to acquire money, I will borrow from a person willing and able to lend me money, at a time when they are available, knowing that I will not repay the debt.

This maxim also fails UT, but it not for logistical reasons. If my lender's house was full of people, F2 determines that I should seek another lender who was more readily available. The contradiction here lies in the fact that if F2 were made common practice, promises would not be believed and so false promising would not be an effective way to acquire money. In this case the reason for F2's failing tells us more about the nature of fraudulent borrowing than F1's logistical problems do.

The Omission of Specifics rule is compatible with Kant's work because the maxims Kant chooses for examples do not include specifics.

In the *Groundwork for the Metaphysics of Morals*, his maxim for fraudulent borrowing resembles F2 much more than F1:

Fk: “If I believe myself to be in pecuniary distress, then I will borrow money and promise to pay it back, although I know this will never happen” (39).

Kant’s version does not include names, times, or places. He even omits the condition that the lender be willing, able and available, but these are pre-conditions to borrowing money so they can be omitted. Kant does not explicitly state anything like the OS rule, but the rule produces maxims similar to the ones he uses in his examples.

In conclusion, false results are a fatal problem for the moral test based on FUL, mainly because they create internal contradictions. Rules governing the creation of maxims are required to solve that problem. One such rule is the Omission of Specifics rule, which dictates that “when deriving a maxim from an action, specifics of the action such as name, date, time and place should be omitted. If any specifics are relevant to the maxim’s end, the conditions that recommended those details should be included.” This rule has been shown to prevent two well-known false results, and it is justified by Kant’s writings about the practical rationality and the nature of maxims. It also prevents maxims from being rejected for trivial logistical problems, and it produces maxims that are similar to those in Kant’s examples. Therefore at least one justified, useable rule exists that can prevent some false results. Possibly, other such rules exist that could collectively eliminate all false positives and negatives.

Appendix: Contradicting popular belief is not a disproof of a moral theory, but it can make a theory unpopular.

Allen W. Wood claims that “FUL and FLN are incapable of disqualifying many obviously immoral maxims” and also seem to forbid some maxims that are obviously morally acceptable (103). He sees this as a disproof of UT, but strictly speaking it is not. I object to the assumption that the moral status of any action (or its maxim) is “obvious,” unless we have already accepted some moral framework. The liveliness of debate in moral philosophy attests to the fact that we have no certain moral standard that can be used to judge others’ belief structures. If Wood means “widely accepted as true” when he says “obvious,” he still has no grounds to criticize UT. In the *Groundwork for the Metaphysics of Morals*, Kant states explicitly that his account of morality is justified by rational principles, not popular appeal (23). The fact that some of its results contradict “obvious” moral standards does not disprove UT.

However, the intuitive repugnance of false results is a practical barrier to acceptance of Kant's theory. In the *Groundwork*, Kant acknowledges that practical examples of a moral principle are good for "procuring entry for it by means of popularity, once [the principle] stands firm"; showing that a moral principle agrees with convention demonstrates "the feasibility of what the law commands" and makes it intuitive (25-6). Kant recognizes that people will understand and accept a moral theory more readily if it can be shown to agree with and justify what they already believe. Kant demonstrates how his different formulations of CI all justify the popular condemnation of fraud, suicide and sloth, probably for this reason. This suggests that a moral theory which contradicts popular belief will be seen as unfeasible and counterintuitive, and will therefore not "procure entry," or be accepted by people as true. Kant desired his theory to be accepted and practiced, so he would certainly not want to oppose popular opinion in trivial cases.

Works Cited

- Kant, Immanuel. *Groundwork for the Metaphysics of Morals*. trans. and ed. Allen W. Wood. New Haven: Yale University Press, 2002.
- Korsgaard, Christine. *Kant's Moral Philosophy*. Cambridge: Cambridge University Press: 1996.
- Wood, Allen W. *Kant's Ethical Thought*. Cambridge: Cambridge University Press, 1999.

Special Relativity and the Time-Lag Argument

JOHN KARDOSH

UNIVERSITY OF ALBERTA

The time-lag argument in the philosophy of perception is often used as an argument against the Direct Realist theory of perception. In short, it is usually argued that because it takes time for the light transmitted (emitted or reflected) from the object of perception to reach the perceiver, perception cannot be of present stages of objects; rather, it must be the case that in perception what we perceive are past, presently non-existent stages of objects. Some Direct Realists, such as George Pitcher (1971) accept the consequences of the time-lag argument but insist that for Direct Realism all that is necessary is that perception is of external physical objects; and nothing for the Direct Realist, Pitcher argues, hangs on perception being of existent stages of those objects. More recently, R.W. Houts (1980) has argued that Pitcher's concession to the time-lag argument has unfavorable consequences for the Direct Realist. Houts claims that if we maintain a Pitcher-type response to the time-lag argument, then it is impossible for us to be spatially related to the objects we perceive. In this paper, I contend that R.W. Houts' argument against Pitcher's response is uninformed. I show that Houts fails to formulate his argument in accordance with the laws of special relativity (STR), and this renders his position implausible. I also display that the STR rule of the 'relativity of simultaneity' has similar consequences in the philosophy of perception to those that arise out of the time-lag argument. I argue that Pitcher's response is the only response available to the Direct Realist grappling with the implications of STR.

In “Some Implications of the Time-Lag Argument,” R.W. Houts argues that George Pitcher’s defense against the time-lag argument can be taken as a *reductio* of the Direct Realist theory of perception. More Specifically, Houts argues that if the Direct Realist is willing to ‘bite the bullet’ and agree that we can only ever perceive¹ “*past* physical events and stages” (Houts, p. 150), then the Direct Realist is forced to admit that the following two propositions are false:

1. At a time, t , we (or our bodies) are at some spatial distance from the events and stages we perceive at t .
2. All the spatially noncontiguous events and stages we perceive at a time are or were at some spatial distance from one another. (p. 156)

Houts, however, insists that it is overwhelmingly plausible that these two propositions are true. Indeed, it does seem reasonable to maintain that we stand at some spatial distance to the external objects we *now* perceive, and that these objects of perception stand at some spatial distance to one another. Nevertheless, as Houts contends, we cannot be at a spatial distance from *stages* of objects that no longer exist, and stages of objects that no longer exist cannot be at a spatial distance from one another.² In a similar regard, one cannot be at a spatial distance from the ‘golden mountain’, and the ‘golden mountain’ cannot be at a spatial distance from the ‘golden valley’, because non-*existent* objects are not spatially located. Houts, therefore, concludes that the Direct Realist response—that we can only ever perceive “*past* physical events and stages” (*ibid*)—to the time-lag argument, reduces the Direct Realist theory of perception to *absurdity*.

Nevertheless, Houts’ argument is incomplete, because it does not account for the implications that the special theory of relativity (STR) has for the time-lag argument. It will be the principle task of this paper to show that a time-lag in perception is unavoidable, given that light travels with a finite constant velocity. In the first part of this paper, I will show that Houts’ argument against Direct Realism loses its intuitive force once we realize that his above two propositions are implausible under the

¹ The term perception in this paper will refer exclusively to visual perception.

² This is not, however, to make the stronger claim that we are never at a spatial distance from *objects* in the physical world, or that objects in the physical world are not at a spatial distance from one another. Rather, the claim is only that, if we accept the implications of the time-lag argument—that it takes time for the light necessary for perception to reach the perceiver—then we can never *perceive* objects at a spatial distance from ourselves, or from one another.

constraint of STR. Houts may be right in saying that Pitcher's response to the time-lag argument is less than desirable, but he fails to realize that Pitcher's response is the only one available to the Direct Realist. The last part of this paper will address some further problems that arise out of STR. In short, I argue that the relativity of simultaneity leads to a type of 'perceptual relativity', where innumerable many, equally valid, reference frames exist in perception; a result of this is that the Direct Realist is cornered into a position where she must make another Pitcher-type response. First, it will be expedient to give a brief exposition of the Direct Realist theory of perception.

Direct Realism in the philosophy of perception is the view that the immediate (or direct) objects of perception are external objects, events and stages.³ That is to say, a Direct Realist holds that in cases of veridical perception, what we perceive immediately are mind-independent external objects. The most radical Direct Realists will also usually maintain that the object of perception must be contemporaneous with the perception of which it is the object (Robinson, 81); if the object of perception is not contemporaneous with the perception of which it is the object, then the object of perception does not correspond to a *present* physical object, event or stage in the external world. The Direct Realist, then, must hold that the object of perception is contemporaneous with the perception of which it is the object, if the Direct Realist wants to hold that the immediate object of perception corresponds to an *existent* object in the physical world.

The time-lag argument undermines the Direct Realist theory by showing that the immediate object of perception cannot be contemporaneous with the perception of which it is the object. In other words, the argument displays that the immediate object of perception cannot be a presently *existent* stage or event in the external world. In its most standard formulation, the time-lag argument is motivated by the fact that light travels with a finite speed ($c = 3 \times 10^8 \text{ m/s}$),⁴ and then shows that because of the finite speed of light it is impossible for the object of perception to be contemporaneous with the perception of which it is the object. Consequently, if the object of perception is not contemporaneous with the perception of which it is the object, then the object of perception cannot be a presently existent external object, event

³ A stage of a physical object can be thought of as an event— a slice out of the history of the object (Houts, 153).

⁴ It is a law of STR that the speed of light is both constant and finite in all reference frames.

or stage. For instance, against Direct Realism, it can be argued that it is an uncontroversial fact that when we perceive celestial bodies, some of those bodies no longer exist; thus, with regards to the perception of astronomical phenomena, in some cases what we perceive are no longer existent objects in the external world (Robinson, p. 80). While the appeal to non-*existent* celestial bodies, in no doubt, strengthens the argument, the true force comes from the fact that, even with regards to objects in our immediate environment, it takes time, how ever short it may be, for the perceiver to receive the light transmitted from the object of perception. So, even with regard to nearby objects, we can never perceive them as they *are* at a present time, but only ever as they *were* some time prior to a present time, because light *always* travels with a finite speed. Therefore, the immediate ‘object of perception’ must be something either non-*existent* or non-physical. The Direct Realist, however, cannot hold the latter,⁵ so she is forced to assert that the immediate object of perception is something presently non-*existent*.

The common response for a Direct Realist to make to the time-lag argument is to bite the bullet and agree that we can only ever perceive *past* physical events and stages—non-*existents*.⁶ George Pitcher articulates this concession well when he states that:

[A Direct Realist] can simply insist that the finite speed of light does not entail that we do not directly see things and states of affairs in the “external world,” but only that we must see them *as they were some time ago*. We see real physical things, properties, and events, all right, but we see them late, that is all. According to a direct realist, it is a mere prejudice of common sense—and one on which the time-lag argument trades—that the events, and the states

⁵ This will not be discussed at length here; however, it should be noted that if the Direct Realist holds that the immediate object of perception is something non-physical— a sense datum, then they have given up Direct Realism.

⁶ Recently, A.D. Smith has used non-existent objects in his response to the problem of hallucinatory perception. In short, he claims that, in cases of perceptual hallucination, what one is aware of is a (Meinongian) non-*existent* object. I think that a similar response to the time-lag argument, as that of Smith’s to the problem of hallucination, is available to the Direct Realist; however, it seems untenable to me that *all* perception is of non-*existent* objects. I lack space to pursue this point further; but for a more informed discussion see A.D. Smith, *The Problem of Perception*, (Harvard University Press. 2002).

of objects, that we see must be simultaneous with our seeing them.
(Pitcher, p. 48)⁷

It is apparent that Pitcher's response in this passage is what fuels Houts' argument. The Direct Realist, Houts asserts, wishes to maintain that a Direct Realist theory of perception can remain coherent and still carry on a Pitcher-type response. That is to say, a Direct Realist can reject the above claim that the immediate object of perception must be contemporaneous with the perception of which it is the object; and instead, she can maintain the weaker claim that the immediate object of perception need only be of an external stage of an object, and not of a present temporal (existent) stage of that object (Suchting, 50). However, while this "weaker claim" may remain consonant with a Direct Realist⁸ theory of perception, Houts argues that such a defense is incompatible with his aforementioned two propositions. This alone, for Houts, warrants a reason to reject a Pitcher-type response as a plausible defense against the time-lag argument.

As mentioned above, in order to perceive a spatial distance, in Houts' view, a perceiver must perceive an object as it *is*; and in order for objects to have spatial distance from one another, they must coexist.⁹ For instance, let's imagine an event (E) where two billiard balls (A and B) strike each other; and let's also imagine that a perceiver (Smith) is watching E. Now, because of the time it takes the light transmitted from E (A and B striking each other) to reach Smith, what Smith sees when he looks at E is a *past* stage (A and B at the *moment* of impact), but A and B at the *present* stage have already collided (let's call this new event E¹). Houts will then argue that Smith, in our thought experiment, is at a spatial distance from E¹, and A and B are at a spatial distance in the E¹ scenario; but Smith does not perceive E¹, Smith perceives E; thus, Smith is not at a spatial distance from what he perceives (a *past* stage-E); and likewise, Smith does not perceive a spatial distance between A and B, because A and B, in his view, are prior to E¹, when in reality A and B are *presently* at E¹. But this conclusion—that propositions one and two are

⁷ For an in depth discussion of Pitcher's response to the time-lag argument, see, G. Pitcher, *A Theory of Perception*, (Perception: Princeton Univ. Press. 1971).

⁸ The radical Direct Realist, sometimes referred to as a Naïve Realist, will not be able to maintain this weaker claim. It is essential to Naïve Realism that the object of perception be contemporaneous with the perception of which it is the object.

⁹ Houts' wants to say that two objects that 'coexist' share a spatial relationship because they exist at the same temporal moment (Houts, 154).

false–Houts claims, is counterintuitive, because when we look at objects and events in the world we tend to think that these objects and events are at a spatial distance from us, as well as from one another. Houts, therefore, concludes that if Pitcher’s is the best response to the time-lag argument, then Direct Realism is an untenable theory of perception.¹⁰

Nevertheless, Houts’ argument loses its intuitive force once we consider the implications of STR. In STR, the notion of “spatial distance” becomes highly relativised, and a new notion of spatio-temporal distance must be introduced. As such, when we describe a relation between a perceiver and an object, or an object and another object, it only makes sense to say that the perceiver occupies a spatio-temporal relation to the objects she perceives, and that objects occupy spatio-temporal relations to one another. So, if we were to describe our above thought experiment again, we would need to say that Smith is at a spatial distance of x (where x signifies a set distance) and a temporal distance of t (where t signifies a set time) from E , and Smith is at a distance¹¹ of $x't'$ from E^1 . Houts makes a point of specifying in his propositions that it is because we cannot perceive events as they *are*, but only ever as they *were* that we have to deny ‘one and two’. While this shows that Houts is, indeed, specifying a type of spatio-temporal relationship, it is not one that fits well into STR. Houts’ confusion lies in trying to separate a temporal relation from a spatial relation: he claims that we can only have a *spatial* relation between objects at t –a present time. However, as I said, there is no such separation in relativity theory. According to STR, perceivers are *spatially-temporally* related to objects, and objects are so related to one another. Thus, when we consider STR, the propositions Houts uses in his argument lose their appeal: a perceiver could only perceive an object as it *is* if she could occupy the same spatio-temporal point as the object of perception, but this is impossible. On a Minkowski diagram¹² we cannot draw two intersecting world lines if there is a *distance* between the objects. Houts’ argument,

¹⁰ For an alternative response to the time-lag argument, see, e.g., W.A. Suchting, “Perception and the Time-Gap Argument,” *Philosophy Quarterly*, pp. 44-56, and H. Robinson, *Perception*, Routledge, 1994.

¹¹ Unless otherwise specified, I will use distance from here on to signify spatio-temporal distance.

¹² In STR, Minkowski diagrams allow for the illustration of the properties of space and time, and for the quantitative understanding of certain phenomena like *time dilation* and *length contraction* without mathematical equations.

therefore, is *absurd* in its own right, and does not constitute a persuasive objection against Direct Realism.

This response to Houts' argument should, nevertheless, inspire us to consider the implications that an STR version of the time-lag argument has on Direct Realism. What remains consistent in the STR version of the time-lag argument with the version stated above is the key premise that light *always* travels with a finite speed. As such, if we consider a single inertial frame as our perceiver (Smith), then it is a consequence of the finite speed of light that Smith will only ever be able to perceive objects, events and stages at a later time than the time the light, that carries the information *necessary* for perception, is transmitted from these objects, events and stages—the minimum lapse being the closest distance a perceiver can be to the worldline of light transmitted by the object of perception. The only way, then, that it would be possible to perceive *objects* in the physical world without a *lapse* in time would be if we could somehow occupy the same spatial-temporal point (at the origin on a Minkowski diagram) as the object of perception, which is not a possibility for numerically distinct objects. It will always be the case that from the *objects'* reference frame, the light transmitted occurs at the *origin*, and from the perceiver's reference frame, the light is received at *xt* (some distance from the origin).¹³ Hence, Pitcher's response, for a Direct Realist, is the only adequate response to the time-lag argument.

It should now be clear that it is merely a consequence of STR that a time-lag in perception is unavoidable. Pitcher's concession acknowledges this point. STR, however, creates other problems for Direct Realism. Up till now we have been considering how the time-lag, in conjunction with the principles of STR, pushes the Direct Realist into a position where they must concede that the object of perception cannot

¹³ It is possible to argue that the temporal present could in fact be extended long enough so that the light transmitted from the object of perception arrives at the perceiver during the same temporal moment. This attempt fails in STR because the notion of the present is a space-time point that is an infinitesimal degree above zero, which means a time-lag would still be prevalent. One might still contend that the spatial-temporal present is bounded to a degree equivalent to that of the speed of light in all inertial frames, making the lag insignificant; for why this cannot be the case, see, e.g., H. Stein, "On Relativity Theory and the Openness of the Future," *Philosophy of Science*, Vol. 58, No. 2 (June, 1991), pp. 147-167. Also, we must consider that the *actual* time-lag will also take into account the time it takes for a percept to be elicited in the perceiver through a series of neural pathways. The time it takes for this biological process to transpire in the perceiver will make the time-lag readily apparent.

be contemporaneous with the perception of which it is the object. This is a problem that occurs for any one perceiver occupying any given reference frame. But if we introduce more than one perceiver into the picture, then we also have to address the ‘perceptual relativity’¹⁴ that results because of the *relativity* of simultaneity.

It is a consequence of STR¹⁵ that the notion of simultaneity is dependent on the reference frame (perceiver). That is to say, it is impossible to say in an *absolute* sense whether two events occur at the same time if they are separated spatial-temporally. What we can say, however, is that if we have two perceivers watching the same object, event or stage, neither will be able to see the object, event or stage bellow the *minimum lapse*. Nevertheless, it is possible for one or both to see the object, event or stage at a *greater lapse*. For instance, if two flashes of light are emitted at two ends of a moving train (stipulating for now that on a Minkowski diagram both flashes are on the time axis at zero, but are at different points on the space axis—one flash at the right end (R), and one at the left end (L)), and there is a perceiver (Smith) in the middle of the train, and another perceiver (Gilbert) on the middle of a platform that the train is traveling towards (the right end being the front of the train), then Smith, being inside the train reference frame at an equal distance from L and R, will say he sees the two flashes simultaneously; however, Gilbert on the other hand, will say that he sees R before he sees L, because, from Gilbert’s reference frame, R has less distance to cover to get to the middle of the train than does L (Kosso, 51-56). Furthermore, also stipulating that both Smith and Gilbert have synchronized *clocks* (i.e., they must count units of time at the same rate, and they must simultaneously indicate the same time), then Smith will perceive that L and R reach him at the same time t^2 , whereas Gilbert will perceive that R reaches Smith at t^1 and that L reaches him at t^3 .¹⁶ This, however, does not mean that Smith sees the event (the simultaneous light flashes) as it *really* is, and Gilbert does not; on the contrary, there is no one objective frame in STR. As a result of there being no objective frame, if everything else is held equal, then the Direct Realist is forced to concede that Smith perceiving both L and R at t^2 is compatible with

¹⁴ By perceptual relativity, I mean the relativity between the time at which objects are perceived in different reference frames.

¹⁵ It is specifically a consequence of the *constant* speed of light in STR that simultaneity becomes relativised.

¹⁶ The times designated here are not important. I only wish to clarify the relativity of simultaneity.

Gilbert's perceiving R at t^1 and L at t^3 ; and a similar conclusion must hold true for innumerable many different reference frames. Moreover, this also entails that while the time-lag will never go below the minimum lapse (i.e., in this case we will say Gilbert's seeing R), it can be greater than this minimum lapse (i.e., Smith's seeing L and R, or even greater still, Gilbert's seeing L). Since, the Direct Realist is willing to make a Pitcher-type concession to the time-lag argument, then they will most likely make the same concession to perceptual *relativity*—the immediate object of perception need only be *of* an external stage of an object. To what extent this “concession” affects Direct Realism is not clear; nevertheless, we have shown that the Direct Realist is forced to make this concession.

In this paper I have presented the time-lag argument from an informed discussion of STR in order to show that Direct Realists are forced to admit that perception is *always* of past physical events and stages—*non-existents*. I have displayed that the force of the time-lag argument relies on the fact that light has a *constant* finite velocity. The *constant* velocity of light from STR has also allowed me to show that it is impossible to go below the minimum-lapse in the time it takes light transmitted from the object of perception to reach the perceiver. Although it is impossible to go below this *minimum*, when more than one perceiver is introduced into the equation, the *relativity* of simultaneity shows that it is possible for the time-lag to increase, and for perceptual relativity to occur between two or more perceivers with synchronized *clocks*. As a result of these implications of STR on the time-lag argument, Houts' argument against Direct Realism loses its intuitive force. In the end, the Direct Realist still has the ability to maintain a Pitcher-type response to the problems that arise when we consider the relation between STR and perception; whether this is a favorable move, is still left open.

Works Cited

- Houts, R.W. “Some Implications of the Time-Lag Argument,” *Philosophy and Phenomenology Research*, 41 (1980), pp. 150-157.
- Kosso, P. *Appearance and Reality: An introduction to the Philosophy of Physics*, Oxford University Press, (1990), pp. 21-93.
- Pitcher, G. *A Theory of Perception*, Princeton: Princeton University Press, 1971.
- Robinson, H. *Perception*. Routledge, 1994.
- Smith, A. D. *The Problem of Perception*, Cambridge, Mass.: Harvard University Press, 2002.

- Stein, H. "On Relativity Theory and the Openness of the Future," *Philosophy of Science*, Vol. 58, No. 2 (June, 1991), pp. 147-167.
- Suchting, W.A. "Perception and the Time-Gap Argument," *Philosophy Quarterly*, 19 (1969), pp. 44-56.

Judging the Judges: Wittgenstein's Sceptical Paradox for Debates in the Philosophy of Neuroscience

WILL ROBBINS

UNIVERSITY OF SASKATCHEWAN

In their work, Philosophical Foundations of Neuroscience, Maxwell Bennett and Peter Hacker argue that modern neuroscientists are labouring under a conceptual confusion— that of wrongly ascribing psychological predicates such as “thinks” or “infers” to parts of the brain, instead of to the whole person. They contend that such use of these folk psychology concepts violates the rules of their use. This argument has met with stiff rebuttals from philosophers such as Daniel Dennett and Paul Churchland, who counter that Bennett and Hacker have overstepped the bounds of their ordinary-language analysis in the case of neuroscientific research. These defenders of the current neuroscience project argue that there are no explicit rules available to point out this supposed violation, and until such rules are made available (and shown to be valid), critics such as Bennett and Hacker have no available justification for such claims against the coherence of neuroscientist’s language use. Within the context of this debate, an examination of Ludwig Wittgenstein’s arguments on the subject of language use and rule following will be undertaken, specifically from the perspective of Saul Kripke’s reading of the Philosophical Investigations as the formulation of a unique Wittgensteinian sceptical paradox. Kripke’s reading of Wittgenstein regards numerical privacy and justification is used to provide a possible rebuttal for Bennett and Hacker to their critics in this debate.

Although Ludwig Wittgenstein's written work has had tremendous influence over the direction of philosophy this past half-century (essentially founding ordinary language philosophy as a school of thought), one area where he has not been particularly well received is the philosophy of neuroscience. While he may be in the margins more often than centre stage, Wittgenstein's backers do occasionally surface to challenge the assumptions in regular use within the philosophy of neuroscience, contesting some of the fundamental conceptual tenants of the mainstream reductionist viewpoint that presently holds sway.

One such challenge to the "brain as causal mechanism" story that has drawn some intense criticism is the authorial partnering of philosopher P.M.S. Hacker and neuroscientist M.R. Bennett (hereafter referred to as B&H). Their *Philosophical Foundations of Neuroscience* is a systematic undermining of the common ascription of psychological attributes and powers of human beings to specific parts of bodies (most often to various parts of the brain). B&H's account of what they see as the conceptual confusions currently overrunning large portions of contemporary neuroscientific research has been met with sharp opposition from philosophers such as Daniel Dennett and Paul Churchland who see nothing wrong (and many things right) with the direction and conclusions of modern neuroscientific research. Dennett and Churchland both dismiss the critique offered by B&H, arguing that any critique of the commonplace ascription of psychological predicates to the brain, at least in the realm of neuroscientific discussions, needs to show how such use goes wrong (something they both think B&H fail to do).

In this paper I will briefly sketch out the argument presented by B&H in *Philosophical Foundations*.... I will then consider some of the objections raised by Dennett and Churchland to this argument and explain why the rather rigid account of Wittgenstein's position espoused by B&H leads them into some difficulties in meeting such objections. Finally, I will examine the notion of Wittgenstein's rule following as presented by Saul Kripke in his *Wittgenstein on Rules and Private Language* as a means of re-evaluating B&H's critique of the use made of folk psychology concepts by neuroscientists to avoid the sting of the Dennett and Churchland objections.

The central theme of the *Philosophical Foundations*... is that much of the current neuroscientific community is in the grip of a conceptual confusion. More specifically, B&H argue that many neuroscientists ascribe folk psychological concepts (such as "believes," "knows," "reasons," "infers," "sees," "perceives," "plans," "maps," etc.) to parts of the human whole, specifically to their brains or parts of their

brains. This move, they argue, results in nonsense, as it only makes sense to use such concepts when speaking of the whole creature (in this case, of the person, not a part of their body). Fairly obviously, this examination of neuroscientific language use (and misuse) is a specific application of the broader Wittgensteinian point about criteria of use. Taking their cue from Wittgenstein's remark §281, "It comes to this: only of a living human being and what resembles (behaves like) a living human being can one say it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious,"¹ B&H attempt to show how this remark cashes out in the realm of contemporary neuroscience as a host of nonsensical ascriptions (i.e., as a category mistake). They dub this faux pas the mereological fallacy (mereology being "the logic of part/whole relations").² This misplaced ascription of psychological predicates is problematic for neuroscience because it leads scientists into a host of possible errors, not the least of which is talking nonsense (or so B&H assert). More concretely, scientists who are busy framing their empirical experimental work with these alleged mistakes are bound to come up empty handed on the far end of their experiments, because there simply isn't any empirical evidence available that could possibly alter the logical-grammatical relations that make up meaning for concepts such as "believes."³ In short, there is no possibility for discovery if scientists are looking for things that could not logically exist, in the same way that no scientific evidence will ever be presented that would force us to reconsider whether there were any married men who were in fact "bachelors."⁴ As B&H say themselves,

...it is not a matter of fact that only human beings and what behave like human beings can be said to be the subject of these

¹ Wittgenstein, Ludwig. *Philosophical Investigations*. p. 97.

² Though obviously related to the homunculus fallacy (that of supposing there is a smaller you inside you doing the watching, thinking etc. that occurs when you engage in such activities) that A.J. Kenny speaks of, B&H adopt this specific terminology because, they argue, it captures more accurately the nature of the mistake being made more generally across neuroscience. See Bennett, M.R., and P.M.S. Hacker, *Philosophical Foundations of Neuroscience*, p. 73. For the Wittgensteinian foundation of this argument see, Wittgenstein, §268.

³ B&H devote a section of their "Methodological Reflections" chapter to explaining why it matters that neuroscience is full of this conceptual confusion. For a more detailed explanation, see, *Philosophical Foundations...*, pp. 408-09.

⁴ The bachelors example is borrowed from, Bennett, Maxwell, and Peter Hacker, "The Conceptual Presuppositions of Cognitive Neuroscience: A Reply to Critics," p. 129.

psychological predicates. If it were then it might indeed be a discovery, recently made by neuroscientists, that brains too, see and hear, think and believe... But what would it be to observe whether a brain sees something—as opposed to observing the brain of a person who sees something?⁵

If we wish to get on with scientific discovery that could take us somewhere, B&H argue, we need to clear the field of such conceptual confusions. To that end, their book takes aim at a whole host of neuroscientist's claims, both past and present.⁶

It is a key element and worth noting that for this account the evidential grounds for the ascription of psychological terms are not inductive ones, rather they are criterial.⁷ B&H's story of neuroscience rife with this conceptual confusion is based on the notion that such psychological predicates are properly assigned to persons on the basis of "logically adequate criteria" concerning their publicly observable behaviour as it unfolds in their physical and social environments. The project for philosophers, they insist, is not to determine what is true or not, but rather, what makes sense to say.⁸ This insistence that their project is antecedent to, and separate from, the empirical considerations of scientists working in the field is one of the things that garners objections from both Dennett and Churchland.

Both Dennett and Churchland object to this account on several fronts, and along the same lines, and therefore, such objections can be treated as constituting a more or less coherent viewpoint, one that I think could reasonably be ascribed to many scientists and philosophers working in this field. This position presents a generally twofold attack on

⁵ *Philosophical Foundations...*, pp. 70-1.

⁶ Though the *Philosophical Foundations...* is a hefty tome of well over 400 pages, it is true (as several critics have pointed out) that the majority of those pages are devoted to a note-by-note description of all the places where the mereological fallacy raises up in the work of contemporary neuroscientists. B&H introduce the notion of the mereological fallacy on page 73 and then consider specific cases (gathered together into chapters on general areas of mistakes, such as "Sensation and Perception" or "The Cognitive Powers") for the next 270 pages. In terms of pressing the philosophical point, it does pretty much begin and end with the parts/whole distinction for B&H, though in their defence, if they are right about this mistake being a mistake, then there is a lot of correcting to be done.

⁷ See *Philosophical Foundations...*, pp. 82, 114-17, 384, as just several examples among many.

⁸ *Philosophical Foundations...*, p. 116.

the notion of the mereological fallacy: firstly, it holds that empirical science does in fact have something to say on the matter of whether such psychological ascriptions as “believes,” “thinks,” “knows,” etc. are rightly attributed to *sub-personal* entities or processes (to borrow a concept from Dennett),⁹ and secondly, it questions the validity of judgements of incoherence from B&H, on a couple of separate grounds.

On the first objection, Churchland argues that in the case of extension of the concept of “map,” B&H’s claim that neural maps are nonsense is “widely known to be false.”¹⁰ Here he presents the empirical evidence for the preservation of a bulls-eye pattern in the neurons on the cortical surface of a monkey exposed to a bulls-eye pattern visually. This, he argues, shows that even though there is no homunculus that looks at this image, the “topographical relationships” are preserved for further “populations of cortical neuron downstream.”¹¹ Thus, though unseen, he argues that they are still “casually effective in steering the behaviour of [an] entire organism whose brain contains them.”¹² This objection is tied into a larger problem he has with B&H on the question of true and false versus correct and senseless. Churchland wants to argue, with his evidence of neural-processes in tow, that the real question is whether the facts back up the hypothesis being made; that is to say, the real questions of neuroscience are not “conceptual” questions for him at all, they are empirical.

Dennett, for his part, contends that no one in neuroscience actually makes the mistake that B&H attribute to them, that really they are engaged in discussions of these folk psychological concepts such as

⁹ Dennett, Daniel, “Philosophy as Naïve Anthropology,” pp. 89-94. It should be noted here that the general agreement on the usefulness of empirical study to these types of questions found between Dennett and Churchland isn’t entirely uniform in the details. For example, Dennett agrees with B&H that pain talk can only be spoken of in “our ordinary way of speaking about pains,” but doesn’t agree that this is the end of the line. His account includes the use of these terms at a sub-personal level using his concept of the intentional stance (where, as I understand his argument, these concepts can be stretched into useful shapes, as long as there isn’t crossover between the sub-personal and personal level). It is beyond the scope of this paper to deal with this theory of Dennett’s and his potential disagreements with Churchland, who, being an eliminative materialist himself, would likely be more tempted to simply throw out the entire language-game of folk psychology and replace it with chemistry of some sort.

¹⁰ Churchland, Paul M., “Cleansing Science,” p. 466.

¹¹ Ibid.

¹² Ibid.

“sees” and “infers” on a sub-personal level, extending their boundaries through analogistic or metaphorical licence in order to tease out a better picture of the causal mechanisms in play when we engage in the multitudes of things we do.¹³ If we are careful not to confuse the sub-personal with the personal level of discourse, then Dennett thinks we can indeed extend these concepts into the realm of neurons and retinal cones.¹⁴

Both of these objections wish to stake out a claim for empirical considerations playing an essential role in solving problems of philosophy of mind here, but both philosophers also use their examples to vault themselves into what I consider the larger and more promising objection. Recognizing perhaps that any specific case of the ascription of

¹³ Dennett addresses the work of several prominent neuroscientists for his concrete examples here, generally focusing on theories that have been alleged by B&H to commit the mereological fallacy. See, Dennett, pp. 87-93.

¹⁴ These two “empirical objections” to B&H rest in part on a conviction that Quine and his naturalist view of the task of philosophy has got it more or less right (which is unsurprising given who Dennett’s thesis supervisor was). B&H, for their part, are not so convinced. Obviously it is beyond the scope of this paper to fully weigh out what such a disagreement over naturalism might come to, but since it informs Dennett’s objections in this particular case, a quick note is warranted. Quine, as I understand him, holds that there is no significant difference between an empirical and a conceptual truth, and that a priori solutions to philosophical problems are foreclosed by his rejection of the analytic/synthetic distinction (see, Quine, W.V.O., *From a Logical Point of View*. Boston: Harvard University Press, 1953.). As it affects the present discussion, Dennett’s objections (and I think, Churchland’s as well) are founded on the notion that empirical research will indeed shed light on the philosophical problems he is interested in solving. While there might very well be many philosophical problems where new empirical evidence could and should change the argument significantly, B&H are right to insist that there is a real difference between conceptual and empirical truths. What seem like trivial examples abound (that no anthropologist will ever discover some new fact about the world that confirms or infirms the proposition “all bachelors are unmarried men” for example), but the larger point is anything but, given that so much of the philosophy of science is enamoured with the idea that they themselves are a key cog in the empirical wheel of new facts. As Hacker himself says, “Conceptual truths delineate the logical space within which facts are located. They determine what makes sense. Consequently, facts can neither confirm nor conflict with them.” Whether such a distinction is trivial or not is obviously the matter at issue between naturalism and ordinary language philosophy, but that is a debate for another time. See, Bennett, Maxwell, and Peter Hacker, “The Conceptual Presuppositions of Cognitive Neuroscience: A Reply to Critics,” p.129.

folk psychology predicates that they might bring forward as an example will be vulnerable to a similar line of criticism from B&H,¹⁵ both Dennett and Churchland move on to argue that B&H have no jurisdiction to make such claims about proper use (though they do this in slightly different ways).

Churchland uses his example of cortical maps to press the objection that B&H are simply being conservative without warrant (the linguistic inertia argument). He complains of B&H that they are being obstinately literal in their interpretations of the use of these folk psychological terms, and that “many readers will find such ostentatiously literal interpretations, of what are plainly meant to be probing metaphors, to be deliberately deflationary and downright uncooperative.”¹⁶ His point dovetails here with Dennett’s, who argues that the claim that folk psychology concepts have rules of correct use that are being violated, “...is question begging. If Hacker were able to *show us* the rules, and show us just how the new uses conflict with them, we might be in a position to agree or disagree with him, but [he] is just making this up.”¹⁷

While their claims of counter experimental evidence seem to come upon the previously articulated response of B&H, this further

¹⁵ Both Churchland and Dennett’s insistence on the validity of the contested use (as “probing metaphor”) fails I think to grasp that their attempts to craft causally significant stories of behaviour out of these metaphors is what is at issue (and not the nuts and bolts of whether it is actually the case that retinal firings can be mapped onto cortical processes). B&H can grant the technical terms; what they are contesting is the larger notion that these folk psychological terms can be rendered out in causal stories. The problem with such use isn’t that there are in fact no neural processes going on; it is a fact about how we use the terms in question, and that story is simply not one of physical or mechanical determinism. We just couldn’t say “Bob signed his rent cheque” and have it make sense if we didn’t presuppose a linguistic story where Bob was a free human who had desires to meet his social obligations, knowledge of our system of commerce, etc.; the story of mechanical causes is simply a completely different-language game. B&H aren’t denying this game its existence; they are pointing out that you cannot poach terms from the more general language game without examining what those terms cash out as. The story of “Bob signed his rent cheque” can never be fully explained by neural firings because it is a story that is circumstance dependant at the level of persons.

¹⁶ Churchland, p. 469. Further to this argument, he parodies their charge of conceptual confusion claiming Newton’s analogy of the moon as being like a “flung stone, constantly falling” would have to be considered ‘nonsense’ by B&H, instead of explanatory (as we mostly take it to be).

¹⁷ Dennett, p. 85.

assertion deserves some consideration. B&H are clearly committed to the notion of internal privacy being a non-starter (as can be seen throughout *Philosophical Foundations...*, especially the chapters on sensations), which is no surprise given their explicit Wittgensteinian bent. But the crux of the objections raised here seems to rest on the notion of who is to count in determining acceptable use of various folk psychological concepts (which is to say, even if Dennett and Churchland were to grant B&H that claims of internal privacy vis a vis sensation-language are senseless, I think they could still press their objections regards who is to count in the community of language-game officials). Put more simply, Dennett and Churchland's objection is a questioning of the legitimacy of B&H's position as members of this particular language-game community, as justification adjudicators.

Wittgenstein himself has an entire section of remarks (the family resemblance section) that aim to show how the lack of rigid definitions for concepts like "games" doesn't prevent us from using them perfectly sensibly. The whole discussion of family resemblance would seem to imply that new use is perfectly possible, an extension of old concepts into new territories (which is what I take Dennett and Churchland to be advocating for in the case of neuroscience).¹⁸ Essentially, this objection cuts to the heart of the notion of rules for new use on Wittgenstein's account of meaning. Since B&H are avowed Wittgensteinians this is an objection that has some traction.

B&H, for their part, don't do much to avoid the sting of such an argument, only briefly touching on the possibility in their final section of the book, "Methodological reflections." There they do concede that concepts can and do change over time, that new applications are incorporated into the existing language-game, but only if the scientific theorists in question are frank and upfront about laying out the new rules for use.¹⁹ This they argue, neuroscientist have not done. But even if we grant them this factual claim, it appears to leave the door open for neuroscientists to revise and become frank and upfront, and surely devoted followers such as B&H wish to echo Ludwig in precluding some extensions of concepts, specifically some of the concepts in question (which obviously cash out in ways relevant to private sensation debates). Further, the supplementary objection raised by Churchland about rule following (that there are simply too many words and concepts in use in everyday language for it to be plausible that we could actually articulate

¹⁸ See, Wittgenstein, §65-84.

¹⁹ *Philosophical Foundations...*, pp. 379-384.

the rules for their use),²⁰ appears to gain a purchase here on the B&H account that it would not ordinarily have (say on Wittgenstein in his own words), since their response rests on explicit codification with regards to these conceptual extensions by neuroscientists.

In this respect I would argue that B&H generally employ a rather more uptight or rigid version of Wittgenstein's account of language usage than he himself would use. And while Wittgenstein's remarks that make up the bulk of what is commonly referred to as the "private language argument" are faithfully rearticulated by B&H in their initial mereological fallacy arguments, there is more to his remarks on rule following that can be brought to bear on these "language-adjudication" worries.

Here I think it is useful to turn to Saul Kripke's account of Wittgenstein on rule following in his work, *Wittgenstein on Rules and Private Language*.²¹ In it, Kripke lays out Wittgenstein's thoughts on rule following as he sees them. He characterizes the private language argument as Wittgenstein examining the more general notions of rule following in language in the special case of sensation-language. For Kripke, the real heart of the matter in the *Philosophical Investigations* is a sceptical paradox that Wittgenstein introduces, namely, that there appears to be no fact about me that I might appeal to that would justify my insistence that I mean what I say I mean (and consequently, that I "mean" anything). This sceptical worry, brought out in his discussion of a "quus" function (that seems to cohere just as well as our regular concept of "plus" in mathematics with regard to our past behaviour,

²⁰ "A normal English speaker has a vocabulary of perhaps 20,000 words, for each of which at least a small handful of rules would be necessary to adequately guide its use. If we make a rough guess of five rules per word, we are looking at a presumptive total of 100,000 rules embraced by any normal English speaker... None of us could even begin to articulate any such mammoth list of rules. Nor is it plausible that our ongoing speech is literally governed by the repeated (on a millisecond time-scale) application of such rules." Churchland, p. 476.

²¹ I am indebted to Tyler Dahl for pointing me in the direction of Kripke for a possible solution here. Tyler has argued in conversation that acceptance of a Kripkian reading of Wittgenstein's arguments as a sceptical paradox might lead to other problems such as a regress to "first rule followers," an interesting notion, but one that we can abandon to him for solution, as what is at issue here (the new extension of folk psychology concepts to parts of the brain) can be adequately dealt with within the context of existing rules as set out by Kripke. For more on 'first rule followers' see, Dahl, Tyler, "Ineffablating the Invention of Clichés: The "Problem" With Wittgenstein's Required Community."

personal intentions etc.) is his way of elucidating a series of remarks in the *Philosophical Investigations* that culminates in §201, "This was our paradox; no course of action could be determined by a rule, because every course of action can be made out to accord with the rule." Wittgenstein goes on to say in that same remark, "What this shows is that there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call 'obeying a rule' and 'going against it' in actual cases."²²

Kripke cashes out this sceptical paradox with his account of Wittgenstein's sceptical solution, namely that we must abandon the individual as a focus of such investigation. As Kripke states, "The situation is very different if we widen our gaze from consideration of the rule follower alone and allow ourselves to consider him as interacting with a wider community. Others will then have justification conditions for attributing correct or incorrect rule following to the subject, and these will *not* be simply that the subject's own authority is unconditionally to be accepted."²³ The point Kripke is making is that justification (a key element of meaning) cannot be a numerically private activity.

Kripke is particularly helpful in that his account brings to the fore a well articulated sceptical argument that I think accounts like Dennett's would be hard pressed to answer. The discussion of "plus"/"quus" in particular draws out not just the impossibility of numerical privacy for meaning; it poses the same sceptical problem to any attempt to find causally necessary meaning in scientific explanation (something that both Churchland and Dennett are explicitly attempting to do). Kripke's discussion of rule following shows us that there can't be anything in a person's head that determines future meaning all by itself. If the clean and precise rules of mathematics are even vulnerable to this scepticism, then what chance does a correlation account of neural firings on top of real world action at the personal level have? The more general problem with attempts to extend these folk psychological concepts is that they, like almost all of the boroughs of our language city, are not conducive to attempts to conflate meaning with determined causation. This is to say; we cannot make sense of them in the new neuroscientific cases as words that carry meaning, unless we refer back to their original community of adjudicators. But the original adjudication that allowed them to mean anything in the first place was one of ascribing intentionality (prefaced on behaviour criteria of various sorts), rather

²² Wittgenstein, §201.

²³ Kripke, p. 89.

specifically at the personal level (which is where intentionality is located in human language-games). And brains, no matter what else they might do, cannot exhibit behavioural criteria on the personal level.

Now while this argument seems to deflate the notion of intentionality terms as empirical explanations, it still appears to leave open Dennett's possibility that neuroscientists are playing a separate language-game, one that B&H are not qualified to adjudicate. After all, he and Churchland already agree (as well as the many neuroscientists who actually use these folk psychology concepts in their explanations of research findings), so they are not numerically private in their rule following. But if we examine the more general point that Kripke is making here, we can see that the explication of Wittgenstein's thought that he provides does in fact preclude the kind of moves that Churchland and Dennett are keen to make. This is because a necessary condition of rule following being numerically public, is that justification is going to rest on surveyable proof.²⁴ The problem here is that the ascription of these psychological concepts to brains can't be accompanied by such proof. Imagine one neuroscientist claiming that a particular cortical axis "knew" what another cluster of neurons was doing further up the line and "reasoned" its way to the next discharge of electrical current, while another disagreed, claiming instead that it "inferred" what was happening and "guessed" when to discharge current. How could either bring forth surveyable criteria that would justify a community of language users preferring one account over the other? What would it come to behaviourally for a brain to "guess" rather than "reason"? As Wittgenstein says, "whatever is going to seem right to me is right... and that only means that here we can't talk about 'right'."²⁵ Not every linguistic realm possibly colonized by a community is habitable it turns out, no matter if you come alone or with friends.

Another consequence of Kripke's account is that we can see that rules are not identical with codified, intentional thought processes. They can't be, since there is no fact about me that could not accord with a different interpretation of the rule in question. The point that Kripke is drawing out is that rules are not "interpretations" at all. When I follow the rule of addition while summing two numbers, I am not laying the rule for following the rule of addition out in my head and then following it. The noteworthy point being just this: given that every rule is variously interpretable, even if Dennett got his way and the "Rules For Ordinary

²⁴ Kripke, p. 106.

²⁵ Wittgenstein, §258.

Language Use” were set out on a stone tablet somewhere, that would not guarantee that each individual reading them could see their force. Wittgenstein is not just spouting snappy one liners when he remarks that every rule is variously interpretable; he is giving us a real insight into the nature of necessary “Form of Life” agreement for meaning to get up off the ground in the first place. His sceptical worry isn't an attempt to show that private language is impossible (that is just one of the consequences); it is an attempt to answer the question “How can we show any language to be possible?”²⁶ Dennett's demand that the proper rules for use of these contested terms be laid out before him must be premised on a notion of rule following as a particular cognitive or mental process, something akin to individual reasoning. But linguistic rule following specifically does not have this characteristic, as can be seen in the discussion of rule following in the *Philosophical Investigations*.²⁷

Thus we can also see that Churchland's worry about navigating 100,000 rules, every millisecond on the millisecond, just to speak ordinary English is similarly misguided. That we are following rules when we use language doesn't mean, in fact cannot mean, that the rules are somehow “in our heads.” As Kripke says, “The relation of meaning and intention to future action is *normative*, not *descriptive*.”²⁸ It simply can't be a deterministic story because we aren't “interpreting” rules at all when we use them. This point about normativity might seem like a banal truism to some (for what else could linguistic meaning be for creatures like us but normative?) but it is a point that I think is missed by those wishing to “discover” the real causal story of human action such as Dennett and Churchland (As Wittgenstein remarks, we often “fail to be struck by what, once seen, is most striking and most powerful”²⁹). Now there may indeed be such a causal story, but it would be difficult to imagine how one might express it in a language-game that was created by creatures like human beings (i.e., the kinds of creatures who can't help but think of themselves as being undetermined, who rely on behavioural criteria all the way through their language-game). It is our “Form of Life” and this of course has implications for what kinds of concepts are possibly grasped by us, in how we have constructed the entire edifice of meaning in communication, in language. This is I think a point never expressly stated by Wittgenstein but brought out nicely in

²⁶ See Kripke, p. 63 for more on this formulation.

²⁷ Wittgenstein, § 80-202.

²⁸ Kripke, p. 37.

²⁹ Wittgenstein, §129.

Kripke's discussion of the sceptical paradox and its solution.³⁰ In this respect, I think Wittgenstein's sceptical paradox about rule following (and consequently, about linguistic meaning) shares much with Kant's attempt to discover the possible conditions of experience, only for Wittgenstein, his attempt is to answer the question, "whence normativity?"

In conclusion, while at first glance the objections raised by Dennett and Churchland appear to back Bennett and Hacker into a Wittgensteinian corner of their own making regards a rule following community, a more careful reading of Wittgenstein's discussion of linguistic rule following as presented by Kripke shows us that such deterministic views about the very nature of rules serves to undermine these objections.

Works Cited

- Bennett, M.R., and P.M.S. Hacker. *Philosophical Foundations of Neuroscience*. Oxford: Blackwell Publishing, 2003.
- Bennett, Maxwell, and Peter Hacker, "The Conceptual Presuppositions of Cognitive Neuroscience: A Reply to Critics," in *Neuroscience & Philosophy: Brain, Mind & Language*. (New York: Columbia University Press, 2007), pp.127-162.
- Churchland, Paul M. "Cleansing Science." *Inquiry: An Interdisciplinary Journal of Philosophy* vol. 48, No. 5 (October, 2005). pp. 464-477.
- Dahl, Tyler. "Ineffablating the Invention of Clichés: The "Problem" With Wittgenstein's Required Community." 313 Seminar paper, University of Saskatchewan, 2007.
- Dennett, Daniel, "Philosophy as Naïve Anthropology," in *Neuroscience & Philosophy: Brain, Mind & Language*. (New York: Columbia University Press, 2007), pp. 73-95.
- Hacker, Peter. *Wittgenstein: Meaning and Mind*. Oxford: Blackwell Publications, 1990.
- Kripke, Saul A.. *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Cambridge: Harvard University Press, 1982.
- Quine, W.V.O. *From a Logical Point of View*. Cambridge: Harvard University Press, 1953.
- Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G.E.M. Anscombe. Oxford: Basil Blackwell, 1958.

³⁰ See for example Kripke at p. 106 where he emphasizes that for Wittgenstein, proof must be surveyable.

Foucault and Freedom: On the Constitutional Efficacy of Practice

ADAM MICHIELS

WILFRID LAURIER UNIVERSITY

Throughout Foucault's work we see the development of two distinct types of subject: the disciplinary subject and the aesthetic subject. The former is constituted entirely by disciplinary practices while the latter plays a role in its own constitution vis-à-vis practices of freedom and care of the self. In this paper I examine Foucault's conception of freedom by providing an account of how a disciplinary subject is formed, how that subject makes a transition to an autonomous, aesthetic subject, and finally, how that subject relates both to itself and to its foundation in an overarching system of power-relations. Through mapping and ordering the processes found throughout Foucault's works, I show with greater clarity the common aspects of the two polarized conceptions of how a subject is constituted and further illuminate the relationship between freedom, the subject and Foucault's idea of power-relations. Drawing on the work of Christoph Menke, I argue that critique is the first autonomous act of a subject and acts as the tether between disciplinary subjectification and the self-constitution of practices of the self. Following from this, I establish a common link between the two subjects, conceptualized as constitutional efficacy, and show how it provides us with the meaning and nature of freedom in Foucault's work: the main practice of freedom is the ongoing process of trying to raise one's constitutional efficacy to an equivalent level of disciplinary practices.

At first blush, Foucault's development of power-knowledge and disciplinary effects seems to lack continuity with his later work on practices of freedom and of the self. Specifically, difficulty arises in trying to see exactly how his theory of practices of the self can be interpreted in the larger context of the whole of his works, including his earlier work on disciplinary systems of power-relations.

On the one hand, Foucault makes remarks suggesting that relations of power and knowledge "invest human bodies and subjugate them by turning them into objects of knowledge"; or that "discipline 'makes' individuals; it is the specific technique of a power that regards individuals both as objects and as instruments of its exercise" (Rabinow, 176, 188). Here we see the subject entirely constituted by the processes and techniques of an all-encompassing model of disciplinary power. On the other hand, Foucault's later work shift focus to a conceptualization of the subject as having its own effects on its constitution, giving us a view of the subject that appears to have a high degree of freedom regarding its make-up. The problem for the subject here is not one of liberation from such disciplinary systems but is a problem of "personal choice, of aesthetics of existence" (Rabinow, 348).

These two approaches look to be mutually exclusive leaving the former to suggest a rather bleak deterministic picture of existence while the latter looks as though it promises individual freedom and an entirely self-created subject. What is more likely, however, is that Foucault was striving to reconcile his early and later works, seeking to more accurately balance the constitutive power of structure and agency as well as of discipline and practices of the self (Gordon, 396). I will address the following questions throughout this paper: what common aspects can be found between these two types of influences on the subject? What balance between structure and agency can be found in Foucault's work? And finally, what can be said about human freedom given such a balance?

The first step in answering these questions is to provide a procedural account of how the subject makes a transition from being a passive product of power-relations to an active, self-creating subject. Such an account would surely be an abstraction away from Foucault's intended reading of his work; things are rarely so linear for Foucault. However through mapping and ordering the processes found throughout Foucault's works, I hope to show with greater clarity the common aspects of the two polarized conceptions of how a subject is constituted and draw some important conclusions as to the relationship between freedom, the subject and Foucault's idea of power-relations. In the second part of this paper I will argue that critique acts as the tether

between disciplinary subjectification and the self-constitution of practices of the self. Following from this I will discuss the nature of power and knowledge with regards to the process of making the transition from one type of subjectification to the other. Finally, I will establish a common link between the two, conceptualized as *constitutional efficacy*, and show how it can answer the three aforementioned questions.

The Procedure of Subjectification

To begin our account, we see the subject in Foucault's early work as entirely constructed by the power employed in social formation (Bevir, 66). The subject is constructed by power which operates on two different levels relative to the subject, the external and the internal. Externally, disciplinary power works through the constant surveillance, supervision and control of individuals, in accordance with some specific concept of normality, thereby establishing the bases of knowledge that are available to an individual.¹ On the other hand, power operates internally through a "confessional technology" that developed within Christianity but managed to extend beyond its confines into the whole of social life (Rabinow, 306). Here, individuals police themselves by examining, confessing and regulating their thoughts, again, according to some specific concept of normality (Bevir, 66).

These operations of supervision and control of the subject, both on the external and internal levels, gain their efficacy through practice. In the context of discipline, "*practice is the equally central technique of a discipline which signifies both 'the subjectification of those who are perceived as objects' as well as 'the objectification of those who are subjectified'*" (Menke, 200; emphasis added). Practice then, as a technique, ensures the constant, protracted effects of that which employs it, in this case, discipline. Conceptualizing discipline as practice allows for its counterpart in Foucault's theories to be introduced – practices of the self.²

Practice is also employed as a technique of care of the self, or as Menke characterizes it, an "aesthetic-existential self-relation" (200).

¹ Here, power is not seen as repressive but as productive. It produces "reality; it produces domains of objects and rituals of truth." This power not only produces the subject but "the knowledge that may be gained of him" (Rabinow, 205).

² For the remainder of this paper I will use Christoph Menke's terms "disciplinary practice" and "aesthetic-existential practice" as well as their counterparts the "disciplinary subject" and the "aesthetic-existential subject." See: (Menke 203).

Here, the constant effect desired from practice is the structuring of our lives as a “problem of personal choice” (Rabinow, 348). So, on the one hand, disciplinary practices “produce subjects who are able to adhere to norms or degrees of normality,” while aesthetic-existential practices “produce subjects who are able to lead autonomous lives or make personal decisions” (Menke 203). Because practice is shared between the disciplinary and aesthetic-existential subjects, they can be considered to be two sides of one coin, a single subject evaluated in different lights (200). However, the pressing questions for this paper are: how does a subject relate to these two types of practices? What kind of mobility between the two, if any, can occur, and how does it happen? The answers to these questions reside in the idea of *critique*.

Criticism for Foucault is a “historical investigation into the events that have led us to constitute ourselves and to recognize ourselves as subjects of what we are doing, thinking, saying” (Rabinow, 46). That is, criticism entails the analysis of the processes and historical instances that subjectified an individual in the first place with the purpose of identifying the possibilities of doing, thinking and saying different things. By identifying these possibilities, the subject can show himself that his situation, while created according to contingent historical events, is not in fact fixed. What this means is that critique “is the movement by which the subject gives himself the right to question truth on its effects of power and question power on its discourses of truth ... Critique would essentially insure the desubjugation of the subject” (Foucault, “What is Critique” 32). It is important to note at this point that critique does not logically imply a difference in beliefs or norms followed by the subject but rather that, as we will see, those rules and norms, regardless of their content, will have been actively, autonomously subscribed to.³

The process of critique, by providing a subject new knowledge regarding possible options for action, or at least knowledge of the *possibility* of other actions, acts as a tether between the disciplinary and aesthetic-existential subjects. On the one hand, critique is the *first* autonomous act and is the first point of departure for any further aesthetic-existential practices. On the other hand, if critique is neglected, the subject must remain as one constituted by disciplinary practices. A

³ Menke argues that aesthetic activities can only be considered successful if they lead to some other outcome than what was decided at the beginning (208). As we will see shortly, all that is required for success is the active employment of the aesthetic-existential practice itself. Success is measured by which type of practice *decided on the outcome*; if it was decided through aesthetic-existential practice, it is a successful aesthetic activity.

transition from the disciplinary subject to the aesthetic-existential subject can therefore only be accomplished in so far as critique is being actively practiced and the knowledge thereby gained is being actively *applied*. However, although autonomous action is motivated and made possible by the new knowledge gained through critical processes, possession of this knowledge by the subject is not enough to apply it in action. Power is also required.

The relationship between power, knowledge and action must be examined further if we hope to gain a better understanding of how both disciplinary and aesthetic-existential subjects are constituted. Christoph Menke provides insight into this relationship with a discussion on theoreticism, voluntarism and free will. He tells us that:

I cannot have control over these activities and performances just as I like, since I only need to practice what I cannot carry out through mere will or decision. The conceptions of the aesthetic-existential and the disciplinary subject thus agree not only in the critique of theoreticism ... but also of voluntarism and its primacy of free will: I can only want what I can do – what I have the power, ability, and the possibility to carry out and toward which or according to which I direct myself (202).

Menke is right to establish the primacy of the will as a central aspect to both the disciplinary and aesthetic-existential subjects since the will, as intentionality, is required for any directed action. Furthermore, because “I can only want what I can do,” power is required of the will as it establishes the degree of possibility of all acts that disciplinary or aesthetic-existential subjects could potentially want to pursue. The conclusion that power is a necessary condition of action, that is, of practice, is valid. However, Menke’s conclusion that the “ability to act ... entails that power comes not only before knowledge; power also comes before freedom” (202), is problematic in that it overlooks the dependence power has on knowledge with regards to action. Action, regardless of the degree of power held by a subject, is impossible without knowledge of that power.

Having the power, ability and possibility to carry out some action necessarily implies that I have the knowledge of those things to begin with. For some action to be possible, I require the knowledge of whether or not I actually possess the necessary power to complete that action. This is not a situation that ends up in infinite regress, debating which comes first, knowledge or power. Rather, it is my contention that for any action to be possible, both the power to complete the action and the knowledge of that power are required to carry it out; knowledge of opportunity, knowledge of the power I possess, and *actually* possessing

the power are all necessary conditions of undertaking any action at all. Power and knowledge thus depend on each other and work in tandem when pursuing possible actions. This conception of the relationship between power, knowledge and action brings us back to Foucault's earlier ideas on these issues, developed in *Discipline and Punish*, where power and knowledge "directly imply one another" (Rabinow, 175). This is power-knowledge and it shares the same characteristics as Foucault originally illustrated: it both creates and limits the possibility of action.

Given that critique desubjugates the subject, autonomous action and the application of new knowledge gained through the critical process is now made possible vis-à-vis the operating of power-knowledge at the individual, subjective level. This allows for the next step in our procedural account: the re-constitution of the subject as an aesthetic-existential one.

By telling us that "there is no power relation without the correlative constitution of a field of knowledge, nor any knowledge that does not presuppose and constitute at the same time power relations" (Rabinow, 175), Foucault forces us to raise the question: with regards to the knowledge gained through critique, the knowledge required for practices of the self, what are the power relations in which the aesthetic-existential subject now stands? Of course the subject still stands in relation to disciplinary practices as they provide the foundation and starting point for any possible transition to an aesthetic-existential subject in the first place. More importantly, however, the practicing aesthetic-existential subject stands in a power relation *to himself*. This subject has the constant, steady goal of accruing knowledge through critique with the intent of exercising power over his very own subjective constitution. This conclusion that the aesthetic-existential subject stands in a power relation to himself shows us the common aspect between the two approaches Foucault follows in his early and later works and gives us the answer to the first question posed at the beginning of this section: what Foucault's disciplinary institutions and the aesthetic-existential subject have in common is the type of effect they have on the constitution of the subject, achieved through their employment of power-knowledge. In other words, the disciplinary and aesthetic-existential subjects can have the same *constitutional efficacy* in their roles of subject constitution.

Constitutional efficacy has two aspects: first, it is achieved through the use of power-knowledge which determines the "forms and possible domains of knowledge" that are available to a subject (Rabinow, 175). Here, power-knowledge is employed through discipline and critique by the disciplinary and aesthetic-existential subjects respectively.

Second, its level of achievement is measured according to a practice's propensity to constitute a subject in its entirety. That is, the degree of constitutional efficacy a practice has depends on its ability to *exclusively* constitute a subject.

Revisiting some aforementioned concepts, we see the disciplinary subject as entirely constituted vis-à-vis external practices such as surveillance and internal practices such as confession and examination. The constitutional efficacy of disciplinary practices is therefore *always maximized* in that they *completely* constitute the disciplinary subject. All individuals begin at this point as they are born into pre-existing structures of power-relations, and the subject formed therein constitutes the precondition to any possibility of an aesthetic-existential subject. This subject is a passive one, created apart from any concept of agency, entirely within the realm of social structure. In contrast to this, the level of constitutional efficacy that can be achieved by aesthetic-existential practices is far more indeterminate.

If disciplinary and aesthetic-existential practices share the common aspect of being able to produce equal levels of constitutional efficacy, this equality is only hypothetically possible. For an equivalent level of constitutional efficacy to occur, the aesthetic-existential subject would have to take every rule, norm, belief and prejudice that he finds himself subscribing to and subject it to the process of critique. If an individual can call into question every facet of their subjective constitution then it would be possible to entirely re-constitute themselves as a new subject. This however, is far too radical an expectation and as such, this level of efficiency is *possible in theory only*; in practice, the aesthetic-existential subject can never achieve the same degree of constitutional efficacy as the disciplinary subject. There are two reasons for this.

First, the task of passing the entirety of one's subjective constitution through the process of critique is impossibly complex and entirely intellectually consuming; second, the specific practices employed by the aesthetic-existential subject cannot be invented. That is, the aesthetic-existential subject cannot create, discover or conceive of these practices on his own; his access to, and choice of, practices of care of the self depends on the circumstances he finds himself in. They are "models that he finds in his culture, and are proposed, suggested and imposed upon him by his culture, his society, and his social group" (Foucault, "The Ethics of the Concern of the Self" 291). This implies a plurality of possible options that become available to the subject upon reflection and as such, the process of selection may require several phases of critical analysis to discover which option is best. In

short, aesthetic-existential practices are imperfect and a subject may need to revisit past issues several times before confidence in his choice of action can be had.

Although disciplinary and aesthetic-existential practices can theoretically have equal levels of constitutional efficacy, in reality these levels will be quite different. However, because they each do in fact have *some* level of constitutional efficacy, we can see both disciplinary and aesthetic-existential practices situated at the same functional level in terms of subjective constitution. This does not simply mean that both practices play the functional role of subject constitution; that is quite obvious. Rather, what is important here is that they constitute subjects using the same methodology: power-knowledge and its processes are passed through a filter, on the one hand discipline and on other hand critique, the output of which is constantly applied over time, in the form of rules and norms, to the subject. Put another way, power-knowledge is employed by the two practices in order to develop outputs that can be applied to the subject. These outputs differ only in content but share the same form. This is the common aspect found in both Foucault's earlier work on discipline and his later work of practices of the self.

A model of constitutional efficacy then allows us to see how a balance between the constitutive power of structure and agency develops in Foucault's work. As I have argued, disciplinary practices always maximize their constitutional efficacy and completely form the disciplinary subject. On the other hand, aesthetic-existential practices have a much more difficult task in attempting to maximize their levels of that same property. It is precisely this inability of aesthetic-existential practices to reach an equivalent level of constitutional efficacy (as disciplinary practices) that allows for the balance between the two. Because the levels of constitutional efficacy cannot be equal, disciplinary practices will *always* play a role in subjectification, regardless of the practicing of the aesthetic-existential subject. Disciplinary practices not only provide the foundation on which aesthetic-existential practices develop, but they have an ongoing presence that both constrains and allows for aesthetic-existential practices. Given this, and the condition that a subject has engaged in the process of critique, the balance in Foucault's theories is found in the subject that is continually constituted by both disciplinary and aesthetic-existential practices simultaneously, over time.

Given this balance that comes about from the interplay between the two types of practice, what can be said about the possibility of freedom in Foucault's work? Generally, freedom in Foucault's work is interpreted as the active participation of the subject in an overarching

system or power relations through practices of the self as practices of freedom. My interpretation is very close to this one. Conscious, reflective activity is a crucial aspect of freedom in Foucault's work. However, my reading of Foucault that led to the above discussion on constitutional efficacy also leads to a further extension of this idea in terms of freedom.

The main practice of freedom is the ongoing process of trying to raise one's constitutional efficacy to an equivalent level of disciplinary practice. Because such equivalency is impossible, there are always further ways that an aesthetic-existential subject can work on themselves, further ways to work on the project of subjective re-constitution. Every change made, every outcome that is decided upon through aesthetic-existential practices, is an elevation in the level of constitutional efficacy of that subject. Although equivalence is unattainable, the disciplinary subject is used as a benchmark for the aesthetic-existential subject. It is the essential point of reference that contains all of the information necessary to begin the process of aesthetic-existential practice.

The methodology followed by practices of freedom looks to be thus: the aesthetic-existential subject looks to the disciplinary subject for content to be critiqued; critique of some aspect of the disciplinary subject is performed; the subject applies the new knowledge gained through critique to himself, changing him, creating, in part, a new subject; the process is repeated. Every time this process is successfully completed, it shows an increase in the level of constitutional efficacy in the aesthetic-existential subject. The practice of freedom then, means having the constant goal of raising one's level of constitutional efficacy to that of disciplinary practices.

Conclusion and Final Remarks

Throughout this paper I have attempted to use a procedural account of how a subject makes the transition from being completely formed by discipline to a subject that is formed autonomously, one that is self-constituted. Through this step-by-step account I set out to answer these questions: what common aspects are there between these two types of subject? How does Foucault establish a balance between the two? And finally, what can be said about human freedom given such a balance?

Critique is the tether between the two types of subject. An individual uses critique to problematize aspects of the disciplinary subject. This practice de-stabilizes the effects of discipline and allows the subject to re-constitute itself in an autonomous manner. This process does not utilize power more so than knowledge, nor does it require knowledge prior to power. Instead, the process of desubjugation and

subjective re-constitution requires power and knowledge to work in tandem as power-knowledge. This allows for a different codification of power relations in two ways: first, the aesthetic-existential subject now stands in a power relation to itself; and second, the forms and domains of knowledge are manipulated, resulting in new kinds of availability of knowledge for the subject.

These processes, together with the caveat that abilities to exclusively constitute a subject will vary between practices, give us a model of constitutional efficacy. This concept provides the answers to the above questions. First, both the disciplinary and aesthetic-existential practices share this property. They both have the ability to constitute the subject. Furthermore, they share the same methodological approach of subjectification where the outputs they develop that are to be applied to the subject differ only in content, not in form.

Following from this, the balance between discipline and autonomy in Foucault's work is found in the inability of the aesthetic-existential practice to match the constitutional efficacy of disciplinary practices. This inability creates a process requiring continual effort where aesthetic-existential and disciplinary practices simultaneously and continuously constitute the subject.

This balance implies a familiar interpretation of Foucault's conception of freedom where an active, reflective disposition is required that is constantly at work, *practicing* freedom. Applying the model of constitution efficacy to this conception of freedom, we see the practicing of freedom as a constant development of one's ability to constitute themselves as a subject. I put it as having the goal of raising one's level of constitutional efficacy to that of disciplinary practice. In real terms though, an agent practicing freedom this way looks slightly different from Foucault's explicit remarks on practices of the self. This agent, while clearly practicing freedom in several ways, takes part in another sort of practice of the self. He must be working not only on the project of freedom, but also on the project of *maximizing his ability* to undergo the processes of critique and self-constitution.

Foucault's philosophical *ethos* was centered on the idea of critique (Rabinow, 45). My assertion that the subject must maximize his ability to undergo processes of critique means that this philosophical ethos must be expanded to include constant deep reflection and contemplation on what critique is, how it should be carried out, and what should be considered deserving of critique. In fact, this expanded ethos can surely be applied beyond critique. In the sphere of daily life, such a philosophical ethos of critique, reflection and contemplation could enrich and expand our conscious experiences in meaningful ways that extend

beyond a constant attempt at escaping the prejudices and attitudes engendered in us by discipline.

Works Cited

- Bevir, Mark. "Foucault and Critique: Deploying Agency Against Autonomy." *Political Theory* 27.1 (1999): 65-84.
- Foucault, Michel. "What is Critique?" *The Politics of Truth*. Eds. Sylvère Lotringer and Lysa Hochroth. New York: Semiotext(e), 1997.
- Foucault, Michel. "The Ethics of the Concern of the Self as a Practice of Freedom." *Essential Works of Foucault*. Vol. 1. New York: The New Press, 1997.
- Gordon, Neve. "Foucault's Subject: An Ontological Reading." *Polity* 31.3 (1999): 395-414.
- Menke, Christoph. "Two Kinds of Practice: On the Relation Between Social Discipline and the Aesthetics of Existence." *Constellations* 10.2 (2003): 199-210.
- Rabinow, Paul, ed. *The Foucault Reader*. New York: Pantheon Books, 1984.

A Metaphorical, Heuristic Account of Aristotle's Doctrine of the Mean

BENJAMIN VISSCHER HOLE IV

LEWIS & CLARK COLLEGE

There are two conflicting interpretations of Aristotle's doctrine of the mean. According to the first, the mean is a metaphorical, heuristic device for finding the right way to feel and act. According to the second, the mean is a quantity. I argue that the text underdetermines both accounts. I further argue that a metaphorical, heuristic account is preferable on the grounds that it is more charitable.

According to Aristotle, virtue is a mean between two extremes.¹ This is commonly referred to as the doctrine of the mean. Aristotle states: [Virtue] is the mean reference to two vices: the one of excess and the other of deficiency. It is moreover a mean because some vices exceed and others fall short of what is required in emotion and in action, whereas virtue finds and chooses the mean. (1107a1-6)

There are two conflicting interpretations of the doctrine. According to the first, Aristotle speaks of the mean metaphorically, as a heuristic device for finding the right way to feel and act. According to the second, Aristotle speaks of the mean literally, as an actual amount of something. Most attention has been over the strength of the latter, quantitative account. Rosalind Hursthouse has argued that this view is false, and that interpreting the doctrine quantitatively attributes a false view to Aristotle.² Howard Curzer has responded to Hursthouse by arguing that Aristotle commits himself to a quantitative account, and that such an account is plausible.³

In this paper, I respond to Curzer's argument and defend a metaphorical, heuristic account of Aristotle's doctrine of the mean. In section one, I briefly explain the traditional, quantitative account and Hursthouse's criticisms of it. In section two, I explain Curzer's response to Hursthouse. In section three, I argue that the text underdetermines both accounts. In section four, I argue that a metaphorical, heuristic account is preferable on the grounds that it is more charitable than a quantitative one. Therefore, I conclude that we should attribute a metaphorical, heuristic account to Aristotle.

I. Hursthouse's Criticisms

J.O. Urmson presents the traditional, quantitative interpretation of Aristotle's doctrine.⁴ According to Urmson, Aristotle holds that that character errs in two opposed ways.⁵ That is, Aristotle accounts for virtue

¹ Aristotle. *The Nicomachean Ethics*. trans. and ed. Martin Ostwald. 1962. (New York and London: Macmillan/Library of the Liberal Arts). For consistency, I substitute "median" for "mean."

² Hursthouse, Rosalind. (1980) "A False Doctrine of the Mean." *Proceedings from the Aristotelian Society*, 81: 57-72.

³ Curzer, Howard J. (1996) "A Defense of Aristotle's Doctrine that Virtue Is a Mean" *Ancient Philosophy*, 16: 129-38.

⁴ Urmson, J.O. (1973) "Aristotle's Doctrine of the Mean," *American Philosophical Quarterly* 10: 223-30.

⁵ Urmson (1973), 225.

as a mean between two extremes. These two extremes are opposed to one another because one is excessive and the other is deficient. Being excessive or deficient is a matter of having too little or too much of something. Having too much broadly includes cases of “on too many occasions” and “too violently”; and having too little includes cases of “on too few occasions” and “too weakly.”⁶ A virtuous person occupies a mean because there is an actual amount of something he or she has which is neither too much nor too little, but just right.

Hursthouse argues that this interpretation is false. First, she argues that character does not err in two opposed ways. Second, she argues that erring with respect to an object is not a matter of having too little or too much of something. Therefore, she concludes that interpreting the doctrine quantitatively attributes a false view to Aristotle.⁷

First, Hursthouse criticizes Aristotle’s view, “One’s character may err in *two opposed ways*.”⁸ She argues that for any virtue there are more than two ways of going wrong. For example, when Aristotle describes being virtuous as hitting a target there seem to be many ways of missing the target.⁹ Hursthouse also argues that Aristotle’s use of excess and deficiency is ambiguous. Even if a virtue only has two ways of erring, the manner in which the extremes are opposed to one another is unclear. For example, there are many ways to say that something is excessive, such as “on too many occasions,” “too much,” “too often,” and “too violently.”¹⁰ Hursthouse thereby rejects Aristotle’s view that character may err in two opposed ways.

Second, Hursthouse argues that the mean between excess and deficiency cannot be identified quantitatively. She offers two counterexamples. The first is the fearless phobic, who fearlessly faces enemies on the battlefield, but is deathly afraid of mice. The second is

⁶ Urmson (1973), 226.

⁷ Although Hursthouse’s (1980) argument is negative, she has recently argued for a “central” doctrine of the mean. This is a metaphorical, heuristic account with respect to *many* ways of going wrong. Instead of finding the mean between two extremes, finding the mean is like hitting a target at the center of a circle. See Hursthouse, Rosalind. (2005) “The Central Doctrine of the Mean”. The Blackwell Guide to Aristotle's Nicomachean Ethics eds. Richard Kraut, (Blackwell): 96-115.

⁸ Hursthouse (1980), 59. Her emphasis.

⁹ See, e.g., 11094a20-25, 1106b31-32.

¹⁰ Hursthouse (1980), 60.

the slim, healthy wicked person, who has a healthy diet but acts wickedly. For example, he steals the food of others. These characters are rightly disposed to feel and act, but wrongly disposed toward objects. Hursthouse argues that these counterexamples show that it is possible for character to err without having literally too much or too few of something. Although these counterexamples are medial, they are also vicious. They are vicious because they go wrong with respect to objects, not because they have the wrong amount of something. The fearless phobic has a phobia of mice, while a courageous person would not. The slim, healthy wicked person desires the food of others, while a temperate person would not. However, erring with respect to an object is not a case of being disposed to too many or too few objects. Therefore, being medial with respect to the right object cannot be captured quantitatively.

In summary, Hursthouse's criticism of Aristotle's doctrine of the mean is twofold. First, she rejects the thesis that one's character errs in two opposed ways. Second, she argues that being rightly disposed toward objects cannot be captured quantitatively. Therefore, she concludes that a quantitative account of the doctrine of the mean is false.

II. Curzer's Argument for a Quantitative Account

Curzer argues that Aristotle is committed to a quantitative account, and that such an account is plausible. He argues that Aristotle holds that virtue consists in having the right quantity of something, and vice consists in having either too much or too little of something. I take his argument as follows:

1. If a quantitative account is a plausible interpretation of Aristotle's doctrine of the mean, then we should accept it over a non-quantitative account.
2. A quantitative account is a plausible interpretation of Aristotle's doctrine of the mean.
3. Therefore, we should accept a quantitative account over a non-quantitative one.

Curzer does not explain the first premise. Perhaps his reason is charity. He states:

[I]f a quantitative doctrine of the mean offers a plausible picture of the virtues rather than a silly picture, then a literal doctrine of the mean is preferable to a metaphorical, heuristic one.¹¹

While Hursthouse argues that a quantitative account is false, Curzer argues that it is plausible. He does not argue that a metaphorical,

¹¹ Curzer (1996), 129.

heuristic account is implausible. Rather, his first premise suggests that a non-quantitative view is somehow not as charitable. Therefore, Curzer likely holds that a quantitative account is more charitable than a metaphorical, heuristic one.

Curzer interprets Aristotle as holding an architectonic consisting in five principles.¹² Curzer's first principle states, "(1) [e]ach virtue is a character trait concerned with a different aspect or sphere of human life."¹³ That is, different virtues apply to different aspects of human life. Whenever Aristotle introduces a new virtue he describes the sphere of human life to which it applies. Since temperance is concerned with bodily pleasures, for example, Aristotle describes its sphere by distinguishing between pleasures of the body and pleasures of the soul (III.10). Curzer's second principle states, "(2) [a] virtue is a disposition for getting all of the parameters right."¹⁴ This is because Aristotle holds that being virtuous entails feeling and acting at the right times, toward the right objects, toward the right people, for the right reasons, in the right manner, and so on (1106b20-24). Parameters are different aspects of feeling and acting that fall under the sphere of a virtue. For example, a person may be intemperate with respect to an amount parameter, by eating too much, or with respect to an occasion parameter, by eating too often. Curzer's third principle is (3) the doctrine of the mean, which holds that virtue is the mean between excess and deficiency (1107a2-6). His fourth principle is an aspect of the doctrine of the mean, "(4) [e]ach virtue is associated with two vices."¹⁵ That is, excess and deficiency are the two ways one may err with respect to each virtue. Curzer presents his fifth principle as follows:

(5) "The right quantity for each parameter is a mean. Each virtue is medial with respect to all relevant parameters; each vice is too little or too much of some parameter(s)."¹⁶

In other words, being virtuous requires having the right quantity with respect to every relevant parameter. For example, a temperate person

¹² According to Curzer, Aristotle's justice is a special virtue that is not structured like the other virtues in *The Nicomachean Ethics*. However, Curzer argues elsewhere that justice is nonetheless consistent with a quantitative account of Aristotle's doctrine of the mean. See Curzer, Howard J. (1995) "Aristotle's Account of the Virtue of Justice" *Aperion* 28: 235-266.

¹³ Curzer (1996), 130.

¹⁴ *Ibid.*

¹⁵ *Ibid.*

¹⁶ *Ibid.*

must neither eat too much, nor too often. By contrast, each vice has the wrong quantity with respect to some parameter. An intemperate person errs toward excess by eating too much, or too often, and errs toward deficiency by eating too little, or too seldom.

Curzer's goal is not only to defend the doctrine of the mean, but also to argue that it is quantitative. He thoroughly discusses principles (4) and (5) because they spell out how the doctrine is quantitative. His discussion of these principles also meets Hursthouse's two main arguments. She criticizes (4) by arguing that there are more than two vices associated with each virtue, and that the manner in which one's character errs is ambiguous. She criticizes (5) by arguing that the mean cannot be identified quantitatively. In order to argue for a quantitative account, Curzer defends (4) and (5).

Curzer defends (4) by arguing that one's character may err in two opposed ways. Aristotle holds that "it is possible to fail in many ways while to succeed is possible in only one way" (1106b21-2). According to Curzer, this is because it is possible to err with respect to any parameter under a single sphere of human action. Again, parameters are different aspects of feeling and acting that fall under the sphere of a virtue. A person may be intemperate with respect to an amount parameter, by eating too much or too little. Alternatively, he may be intemperate with respect to an occasion parameter, by eating too often or too seldom. Even under a single sphere of human action it is possible to err in many ways. However, this is consistent with Aristotle's view that there are two opposed ways of erring. Eating too much and too often are both excessive with respect to temperance. By contrast, eating too little and too seldom are both deficient. While there are many ways to go wrong, every way errs either toward excess or deficiency. For example, Aristotle's account of courage spells out multiple parameters. He states that being courageous requires getting all of these parameters right:

Accordingly, he is courageous who endures and fears the right things, for the right motive, in the right manner, and at the right time, who displays confidence in a similar way. (1115b18-19)

There are many ways not to be courageous, but only one way to be courageous. Getting any parameter wrong errs toward excess or deficiency. A person errs toward cowardice by exhibiting too much fear or by exhibiting fear too often. Fearing too much involves an emotion parameter and fearing too often involves an occasion parameter. While the cowardly person may err with respect to either parameter, he nonetheless errs toward cowardice. Similarly, a person may err toward rashness by exhibiting too much confidence or by exhibiting confidence too often. Erring with respect to either parameter is rash. Once different

parameters are considered, erring in many ways is consistent with erring in two opposed ways.

Curzer defends (5) by responding to Hursthouse's argument that "*right object* and *right reason* [and *right amount*]... cannot be specified as means."¹⁷ Her counterexamples of the fearless phobic and the slim, healthy wicked person are characters that are rightly disposed to feel and act, but wrongly disposed toward objects. According to Hursthouse, being wrongly disposed toward an object is not a case of possessing too little or too much of something. Curzer argues that Hursthouse conflates object parameters with other parameters. The fearless phobic is not vicious because he fears the wrong objects, viz. mice. He is vicious because he fears mice too much.¹⁸ The slim, healthy wicked person is not vicious because he is wrongly disposed to the wrong objects, viz. the food of others. He is vicious because he desires the food of others is too much. Therefore, Hursthouse's counterexamples are not rightly disposed to feel and act.

Curzer thereby argues that Hursthouse's counterexamples of the fearless phobic and the slim, healthy wicked person are not counterexamples to (5). They are, at best, counterexamples to a stronger principle:

(5') "[I]f a person is vicious in respect to some parameter, then he or she goes to excess or defect with respect to that parameter."¹⁹

If a person is vicious with respect to an object parameter, (5') holds that he goes to excess or defect with respect to an object parameter. However, (5) holds that he may go to excess or defect with respect to a different parameter. According to Curzer, Hursthouse commits Aristotle to a stronger thesis than necessary. Aristotle need not hold (5') because being wrongly disposed toward objects may be explained through excess or deficiency with respect to a different parameter.

¹⁷ Curzer (1996), 133. Curzer adds the emphasis and the bracketed "[and *right amount*]" when referencing Hursthouse's argument.

¹⁸ Curzer also argues that the fearless phobic would not be vicious in the first place. Aristotle states that the fear of mice is "brutish cowardice" (1149a7-8). So, a fearless phobic would "exhibit mental illness, not moral failing." See Curzer (1996), 136.

¹⁹ Curzer questions whether Hursthouse's counterexamples are even counterexample to (5'), but holds that counterexamples to (5') are possible. See Curzer (1996), 133.

In summary, Curzer responds to Hursthouse's main arguments. He explains that virtue may err in two opposed ways vis-à-vis different parameters. He further argues that Hursthouse conflates parameters. Therefore, he explains how a quantitative account of Aristotle's doctrine of the mean is plausible.

III. Does the Text Support One Account Over the Other?

While Curzer explains that a quantitative account is a plausible one, the text is not decisive. He concedes this:

Now a metaphorical interpretation cannot be defeated by even the most painstaking textual analysis, for every passage couched in quantitative terms is neutralized in advance by suggestions that the quantitative terms are really metaphorical.²⁰

In the text, there is a gap between the quantitative terms and the doctrine of the mean. The strongest suggestion that the doctrine of mean is quantitative occurs when Aristotle discusses the mathematical mean. Aristotle states:

The "equal" part is something mean between excess and deficiency. By the mean of the entity I understand a point equidistant from both extremes. (1106a29-31)

In this passage, Aristotle unquestionably uses quantitative terms. The quantitative terms, however, do not apply directly to the doctrine of the mean. Aristotle transitions from the mathematical mean to virtue through several steps. First, he quantifies the distance between the mean and extremes with numbers (1106a33-35). He then compares the quantification of numbers to the quantification of pounds of food, and considers how much food Milo the wrestler should eat (1106a35-1106b5). Next, Aristotle discusses how every *techne*, such as physical training, aims at a mean (1106b5-15).²¹ Aristotle concludes "that virtue aims at a mean" (1106b15) because virtue is better and more precise than any *techne*. Since the discussion of the mathematical mean is several steps removed from virtue, the doctrine is not explicitly quantitative.

Curzer argues that Aristotle commits himself to a quantitative account, and that such an account may be found in particular virtues. He cites passages from Aristotle's discussion of courage and temperance that

²⁰ Curzer (1996), 129.

²¹ A *techne* is a skill or craft.

make (1), (2), (3), and (4) explicit.²² He then states that “[t]hese passages also indicate” that Aristotle holds (5).²³ However, (5) is not, in fact, explicit in Aristotle’s discussion of courage and temperance. Curzer infers that Aristotle holds (5) from the fact that he holds the first four principles. He may regard (5) as a plausible consequence of the first four principles. Yet, it is not clear that Aristotle must hold (5). Indeed, Aristotle also makes the first four principles explicit:

- (1a) “[T]he only remaining alternative is that they are characteristics. So much for the genus of virtue. It is not sufficient, however, merely to define virtue in general terms as a characteristic: we must also specify what kind of characteristic it is ...” (1106a10-15). Whenever Aristotle introduces a new virtue, he sketches the sphere of human action it falls under. For example, Aristotle devotes a chapter (III.10) to explaining the sphere of temperance.
- (2a) “But to experience all of this at the right time, toward the right objects, toward the right people, for the right reason, and in the right manner – that is the mean and the best course, the course that is the mark of virtue.” (1106b20-24)
- (3a) “[Virtue] is the mean reference to two vices: the one of excess and the other of deficiency. It is moreover a mean because some vices exceed and others fall short of what is required in emotion and in action, whereas virtue finds and chooses the mean.” (1107a1-6)
- (4a) “[Virtue] is the mean reference to two vices: the one of excess and the other of deficiency.” (1107a1-2)

Aristotle clearly holds that parameters exist, that virtue requires being well disposed toward all relevant parameters, and conversely, that vice requires being wrongly disposed toward any relevant parameter.

However, it is not clear that the mean is the right quantity with respect to all relevant parameters. The first four principles do not entail that the relationship between the mean and the extremes is quantitative. Besides quantitative terms, Curzer’s (5) offers nothing that is not already in the first four principles. Only some non-quantitative aspect of (5) may

²² See Curzer (1996), 131-132. Courage: (1) 1115a10-24, 1117a32-33, 1117b7-9; (2) 1115b15-19; (3) 1117a1-17; (4) 1115b24-1116a2, 1116a4-7. Temperance: (1) 1118b8-33; (2) 1119b16-18; (3) 1118b21-27; (4) 1118b27-31, 1119a5-6.

²³ Curzer (1996), 132.

be extrapolated from these four principles. (5') may be revised into a weaker principle that allows different parameters:

(5a) If a person is vicious in respect to some parameter, then he or she goes to excess or defect with respect to *some* parameter.

(5a) is explicit in the text, but it does not suggest that the doctrine of the mean is quantitative. Since the text underdetermines both accounts, it is plausible to think that Aristotle holds a quantitative one. However, there is simply no evidence to commit him to one.

IV. An Argument for a Metaphorical, Heuristic Account

I argue that we should accept a metaphorical, heuristic account over a quantitative one. My argument goes as follows:

1. If the text supports two conflicting interpretations, we should accept the more charitable one.
2. The text supports two conflicting interpretations of Aristotle's doctrine of the mean: a metaphorical, heuristic account, and a quantitative one.
3. A metaphorical, heuristic account is more charitable than a quantitative one.
4. Therefore, we should accept a metaphorical, heuristic account over a quantitative one.

I have explained that there are two conflicting interpretations of Aristotle's doctrine of the mean, and that the text does not help us decide which account to accept. For that reason, I appeal to the principle of charity, which asks the interpreter to "maximize the truth or rationality of the subject's sayings."²⁴ I argue that a metaphorical, heuristic account is more charitable for three reasons.

First, it is a *prima facie* violation of the principle of charity to commit Aristotle to a stronger thesis than necessary. There is no textual evidence that Aristotle holds (5), only that he holds (5a). Therefore, Curzer is guilty of the same argument he uses against Hursthouse, that she holds Aristotle to (5') instead of (5).

Second, Aristotle's discussion of *techne* suggests that the quantitative terms are metaphorical. This is because a craftsman's perfect piece of work cannot be captured quantitatively. When transitioning from the mathematical mean to the doctrine of the mean, Aristotle explains how every *techne* aims at some mean.

²⁴ "The Principle of charity". Blackburn, Simon, ed. 1994. The Oxford Dictionary of Philosophy. (London OUP): 64.

[T]his is the reason why it is usually said of a successful piece of work that it is impossible to detract from it or add to it, the implication being that excess and deficiency destroy success while the mean safeguards it (good craftsman, we say, look toward this standard in the performance of their work). (1106b9-14)

Milo the wrestler may use the mean heuristically, for example, imagining a metaphorical mean. Quantitative terms may be a tool for the imagination. In the case of *techne*, quantitative terms are a method for getting all of the parameters right, by experimenting and evaluating possible outcomes. For example, one learns how to be a soldier, horse-trainer, cobbler, and so on, through practice. Each of these examples, military skills, training horses, and shoe-making, points toward a metaphorical mean. That is, aims such as military actions, trained horses, and shoes cannot be captured quantitatively. A craftsman aims to produce good products; he does not aim toward an actual amount between two extremes. Therefore, Aristotle's discussion of *techne* suggests that aiming in this way is a heuristic process that aims toward an unquantifiable mean.

Finally, unnecessarily committing Aristotle to (5) results in an absurdly rigorous set of parameters. The quantitative account stipulates that the mean of each parameter is identified in terms of quantity. According to Aristotle, a quantitative mean is quantitatively exact because it is equidistant from both extremes (1106a33-35). For example, it would be ridiculously demanding for Milo the wrestler to meet amount, occasion and way parameters. Milo would have to eat a quantitatively exact amount of food, to be consumed on the quantitatively exact occasions, in a quantitatively exact way, and so on. If Milo is even a fraction of an ounce off in the food he consumes, then he is vicious with respect to an amount parameter. If Milo consumes his food even one second too early, or one time too often, then he is vicious with respect to an occasion parameter. If Milo consumes his food even one second too quickly, then he is vicious with respect to a way parameter. The result is absurd, which is an uncharitable account of Aristotle's doctrine of the mean.

V. Conclusion

There are two conflicting interpretations of Aristotle's doctrine of the mean. According to the first, the mean is a metaphorical, heuristic device for finding the right way to feel and act. According to the second, the mean is a quantity. Hursthouse has argued against the latter interpretation; Curzer has defended it. However, if we only look to the text for help, we are left at a standstill. For that reason, I look to the

principle of charity for help. I have argued that a metaphorical, heuristic account is more charitable than a quantitative one. Therefore, I conclude that we should attribute a metaphorical, heuristic account to Aristotle.²⁵

²⁵ I am most grateful to my advisors, Joel Martinez and Nicholas Smith, for their helpful comments on earlier drafts of this article.

On Late-Capitalist Desire

BRENT VIZEAU

UNIVERSITY OF ALBERTA

Contemporary capitalism, in having overcome the limitations on production once faced, now faces a limited capacity for consumption. Because economic models demand growth, producers have to find ways to increase consumerism. I will argue that increasing market demand is a matter of creating desires in individuals for produced goods, and that the work of postmodernist Jean Baudrillard together with that of psychoanalyst Jacques Lacan provides a compelling lens for seeing how this is taking place. Baudrillard's notion of the hyperreal as a condition where we model our lives on copies of simulations of reality that have no real basis, along with an understanding of Lacanian desire formation as the mechanics of how we end up desiring the simulacra of Baudrillard's hyperreality, provides insight into how desires can be manipulated, and how and why advertising and media control can be so effective.

French postmodernist philosopher Jean Baudrillard and French Psychoanalyst Jacques Lacan together provide one lens for unpacking the logic of late-capitalism. Virtually unconstrained by limitations of productive ability, capitalism is now primarily constrained by the limitations on the amount we can consume. In order to overcome the limited amount of consumption, producers have to find ways to create a greater demand for their products, and for consumer goods on the whole. Baudrillard's concepts of "simulacra" and "hyperreality" are an effective means of understanding the artificial creation of demand from consumers through the manipulation of what we take to be real in the symbolic economy. The Baudrillardian concepts provide an understanding of how the capitalist "They" presents the public with artificial images of "reality," which occupy the position of Lacan's "big-Other." Understanding desire formation in the Lacanian sense then allows us to see how people might be unwittingly manipulated into further consumerism to meet the need, and satisfy the logic, of late-capitalism.

First I will sketch briefly what I mean by the "logic of late-capitalism." We could begin by looking earlier yet, however, it should suffice to start by looking at early industrialism. In the early years of the rationalized production of goods (and this is my own description of basic economics) practices of production were aimed continuously at more efficient production so that production could approximate or approach demand, thereby maximizing the purchase of products, resulting in the maximization of profits. The aim to overcome productive limitations can be called the "logic of early-capitalism." Early-capitalist economics was driven by demand but limited by productive capacity in many ways, though likely not for every product.

Conversely, however, we find that in contemporary capitalist economics, productive capacity has, in many ways, exceeded the demand for producible goods (given market definition, etc...). Technology and the replacement of human labor with machine/robot labor, along with other improvements in productive efficiency, have led to our being able to produce as many of any given widget as the market might demand, and, in most cases, far more. Again, there are some instances where this does not hold for one reason or another, as, for instance, in the case of a limited amount of a specific raw material needed to produce the good. But, for the most part, and in most ways, we have overcome our limited production capacity, and have achieved the ability to supply the demand of the market. But, because capitalist economic models are based on perpetual growth, resting production at the demand of the market is taken to be negative and what can be called "market stagnation." In short, late-capitalism seeks ever greater profits and ever greater growth and so

needs to sell an ever greater quantity of their goods (or services); yet if producers can produce up to the demand of the market, there exists a market or growth threshold, limiting the sought after “ever greater” profits. So, in late-capitalism, the logic that drives the machine, I argue, is the aim to overcome consumer limitations, rather than limitations of production. The producers, and service providers, must find ways to create a demand for their products in order to satisfy their pursuit of constant growth. I want to show how we might understand this creation of demand, which really amounts to creating desires in the consumers, through the work of Baudrillard and Lacan. There are other lenses through which to read culture and explain this same phenomenon, and it seems likely that given the complexity of our social reality, no lens is able to see it all. My hope is that through these thinkers, one might gain insight into the nature of consumer desires in our late-capitalist era.

The two concepts central to the thought of Baudrillard, and central for the purposes of this paper, are the “simulacrum” and the “hyperreal”; understanding these interrelated concepts is crucial to understanding his description and critique of contemporary western culture. The first, and central concept for Baudrillard is the concept of the simulacrum (plural: simulacra). A simulacrum is effectively a copy of an original that does not exist (Baudrillard, “Simulacra and Simulation” 6). He traces the progression of simulacra through three stages (Baudrillard, “Selected Writings” 166-184): the first order simulacra are representations (masking and perverting) of what I call “organic” reality.¹ In a sense this order of simulacra is one of counterfeiting, for example, a painting of a real tree as capturing a presence. The second order of simulacra, that of masking the absence of reality, can be entailed by the mass production of replicas, or the reproducibility, of reality. Here, each of the mass of copies is just as real as the prototype, and the original copy becomes indistinguishable from any further copy, for example, the first print of a famous painting is perfectly (re)produced *en masse* for anyone to own.

Most interesting, and most important, is the third order of simulacra. When the term “simulacrum” is employed, this is generally the order that is being referred to. In this order we lose our grip on reality. The “real” takes on a new meaning, and we have the ushering in of the “hyperreal.” The third order simulacrum is a copy with no original. Simulations either precede their “real-world” occurrences, or occur as

¹ I use the marker “organic” to distinguish what we normally take to be reality from hyperreality, which I will discuss further below.

representing reality but with no basis in reality (Mann 287b). For an example of the former case, a representation before the event, we can turn to movies or science-fiction novels that predict events that follow after their production. Douglas Mann gives the example of the film *The China Syndrome*, about a nuclear reactor catastrophe that preceded the incident at Three Mile Island. The latter case, a copy with no original, is the basis for hyperreality, and can be used to characterize our contemporary era, that of late-capitalism (Mann 288a). A straightforward example of this kind of simulacrum is the family sitcom on American cable television. It is a fiction that is supposed to exemplify, that is, it presents itself as, the typical American family, but the “real” typical American family is nowhere to be found as the basis of the programming. Once a sitcom of this type finds expression on a medium as widespread as television, it comes to have a normalizing function. The family sitcom, a simulation of a reality that does not exist, comes to serve as a model for American families. This is the state of hyperreality that Baudrillard describes. Of course, the obvious question that arises in any description of a third-order simulacrum is whether any such real case exists? It might be the case that some family happened to fit the model of the sitcom prior to the sitcom, but the sitcom is not based on that case, nor was that family “typical” prior to the influence of that sitcom on “real” American families. Baudrillard’s concepts are not strict and exclusive, rather they point in directions that actual sociality tends.

The influence of simulacra on our lives results (or could result) in a state of hyperreality. Hyperreality is where reality simulates a simulacrum, or where “real life” begins to model a copy with no original. In a sense, cultural symbols become completely detached from “reality” and serve to constitute their own hyperreality (Baudrillard, “Simulacra and Simulations” 1). This state is a latent goal of the advertising industry that drives consumerism, the engine of late-capitalism. As I have argued, this brand of capitalism is no longer constrained by limited productive capacity, as we the industrially/technologically-advanced western-world have overcome such restrictions, and are constrained only by the capacity to consume. Producers have to create a market for their goods to be consumed, and so they have to create desires in individual buyers (Mann 286a), which is why Baudrillard says, “today this ‘material’ production is that of the hyperreal itself” (“Simulacra and Simulations” 23). It is easy to be skeptical of just how powerful advertisers, media, and the moguls of popular culture can be at manipulating desires. After all, my desires are my most intimately “me” features aren’t they? I will use an example to demonstrate how an entertainment industry has created a desire in people via hyperreality, and only point to some of the

cultural issues that have manifest. One should see the examples I use as illustrative of the ways in which Baudrillard's concepts could play out. Of course some of the details might be contestable, but I think that they are still useful for making the theoretical point.

The hugely profitable pornography industry has been able to attract vast audiences, and along with them, vast amounts of "consumption" with respect to the images made available. This is an industry easy for critics to attack for any number of ethical issues, from coercion, to blatant sexism, to drug and alcohol dependency among the actors, to various other reasons. I will not go further into these problems, nor will I explain in any depth how I see the genealogy of the industry from the current situation. My speculation is that it is likely that profiteers have pushed the envelope trying to provide images not available elsewhere, creating a demand for more and more extreme images, and an image of sex that no longer holds a great resemblance with where it started, save for the basics. I want, instead, to look at a common scene from the existing vast quantity of pornographic material available—one that I see informing the "real" sex lives of my peers.

Since the bulk of pornography is marketed to younger heterosexual men (a demographic that seems to be something like late teen years to middle age), who serve as the majority of viewership, I will describe this example as the simulation of the desirable sexual encounter for younger men. Obviously there are other genres of porn, but I feel as though this is possibly the most lucid and prevalent example. Encounters with the blonde, overly made-up, plastically enhanced female porn star have become symbolic of a desirable sexual encounter. The porn star, free of cellulite, body hair or odor (something lost to pornography), is an exemplar of the desirable woman. An emerging trend in pornography, not that it is has not existed long before the period of late-capitalism, only that it is now becoming "normal" with respect to the ideal sexual encounter for young men, is the act of aggressive, if not violent, anal penetration of the female. The typical scene likely has an abnormally well endowed man, beyond desirable for the anatomy of most women, dominating and degrading a woman who, though in obvious discomfort, begs for her treatment as though she needs it. This kind of sexual imagery is informing the tastes of men, who seek out the acceptance of this behavior from their partners, resulting in the pressure for women to not only model their appearances, to some degree, on that of the porn star, but also their behavior to satisfy men.

I give the details of this kind of pornographic scene to demonstrate how it is a simulation of the typical desirable sexual encounter for younger men (or at least a growing portion of younger men

with access to the internet), but that it is a copy with no original. First, there was likely no such *typical* desire before there was a scene depicting it; and second, the woman, begging to be forcefully and painfully anally penetrated by such a grossly exaggerated penis, is hardly likely to be an element of an original, for which this is the simulation. In this sense, the scene is a simulacrum. Moreover, since this simulacrum is now acting as the model for gratifying sex among the viewership, it is changing the “reality” of sexual encounters; we can say that this pornography has created a state of hyperreality. The desires that are the most “me-like” features of myself, organically my own, are now the product of an artificial reality. People often make similar critiques in terms of ideology, and if we were to put this point in those terms, it would be like saying that the ideology gets inside us, in the most intimate ways. The real of the organic is lost and replaced by hyperreality, the new reality of symbols without ties to anything less than other arbitrary creations. Symbols no longer represent the world-out-there, but other symbols alone (copies of a copy with no tie to the world-out-there).

Another fairly obvious (I think) question arises; that is, is this kind of example really a description of all of reality, or just the “reality” of some? Since not all people watch the family sitcom or this kind of pornography, can we say that ours is a hyperreality? On the one hand, Baudrillard should be taken as warning us about what can happen if this type of example becomes prevalent enough. There may be examples, either actual or imagined as possibilities, which are so widespread as to usher in an actual hyperreality. On the other hand, we can read Baudrillard as critiquing our culture as one where we are bombarded by these kinds of examples, so that even if I do not watch television or pornography, I will be overwhelmed by other images and influences to the point that I lose hold of, what I call, organic reality, and take our hyperreal state to be that organic reality.

Although Baudrillard’s description is useful, there is, however, a gap in our Baudrillardian description of how desire can be manipulated, when taken by itself. To this point we see that cultural symbols like the sitcom are copies with no original, and that ours becomes a hyperreality when our lives begin to model those simulacra. What is missing in this account is how we go from exposure to simulacra to desiring and modeling our lives after them, which we would then call hyperreality. I will have to say something about how desire formation works in order to close gap in the account and make full use of what Baudrillard offers. For this, I turn to the philosophical psychoanalysis of Jacques Lacan.

How it is that desire gets cashed out, like nearly everything in Lacan, is slippery and changing from one part of his work to another. I

do not intend to get into the nuances of the evolution of his concepts; rather I want to give as straightforward an account of desire from Lacan as I can. The most direct formation that Lacan gives, that “Man’s desire is the desire of the Other” (“Seminar XI” 235), is the one I will focus on here. He extends on this formulation slightly in the *Écrits*, saying of man’s desire, “Its object is a desire, that of other people, in the sense that man has no object that is constituted for his desire without some mediation” (148). For Lacan, desire is always either the desire of what the other desires, or the desire to be the object of the other’s desire (Evans 37-8). In this sense, desire is always formed in a social context. An example of the first notion of desire is when I see people who occupy symbolically valued positions in possession of, for instance, a particular brand of clothing; I recognize what the other has, by virtue of their own desire, and desire it for myself. This can also work in terms of seeing others *en masse* desiring a particular item², which none of them need possess, so long as I perceive that they desire it. In perceiving the mass desire for the item, I desire it for myself so that I might be recognized as the one who fulfilled the desire that the many had for themselves. Object desire aims at recognition from the other, which is again the aim, though more directly, in the second notion of desire. The second notion of desire, to be the object of the other’s desire, would be fulfilled in achieving for myself the place of their desire. If I was to secure the sought after item, I would be the “item-bearer” that the other wishes to be themselves. I, “the one with the item,” would then be the object of the other’s desire, since they would desire to be “the one with the item.” This notion of desire is the desire for recognition in a more direct way than the first, though each is similar. Again, the desire of the other is either to share the other’s desire, or to desire to be the object of the other’s desire. These are different ways that desires can be formed, but they both hinge on the desire of the other.

Now, to complicate the case further, there need not be a concrete “other” whose desire I take on as my own. For Lacan, there is a symbolic position called the “big-Other”, of which Dylan Evans says, “The Big Other designates radical alterity, an other-ness which transcends the illusory otherness of the imaginary because it cannot be assimilated through identification” (133). The big-Other functions in several ways, and it can often equate, as it will for our purposes here, to the “They,” the

² Sharing the desire of the other need not be object desire. I use this type of desire because it is simple and lends itself to the conversation about capitalism and consumption.

“Someone,” or some other similar construction found elsewhere variously. For example, when someone says “people always look out for themselves first,” they are not speaking about anyone in particular, and it might be that no such people exist, only that there is this generic placeholder of “the one who is supposed to look out for his/herself first.” In Lacanese, the phrase often used is “the subject supposed to know” (Žižek 27-29), when, for example, I do not know and you do not know, but we say that “they” or “people” or “someone” knows. In our reading of late-capitalist culture, it is effective to see the big-Other as the symbol for “the one who desires,” in which case, my desire can take on the nature of either “the one who desires,” the big-Other, or the object of the desire of the big-Other, where neither case has a definite “other,” the desire of whom, is the root of my own desire.

Since my desire is a matter of sharing the desire of the other, the introduction of the big-Other introduces also the possibility of sharing the desire of no one in particular, but with the generic placeholder of “the one who desires.” This makes sense of why, after watching the simulacra of the family sitcom or the pornographic scene, my desire is affected in a way that manifests in me modeling my life, or aspects of, after the simulacra. Sharing the desire of what we perceive to be others “out there,” perhaps many, perhaps successful, or cool, others, but no other in particular, is something that is quite actual, and so perhaps Baudrillard’s hyperreality, too, is actual in the same sense.

Baudrillard gives us a way to understand how an advertising company, for instance, could create an image that presents itself as typically desirable, that is, that people out there have it, and enjoy it, and want it, etc.; Lacan offers a way to understand why this affects our desires: because we desire what the other desires, and in the case of the simulated other of advertising’s simulacra, what the big-Other desires; and Baudrillard again gives a description of this final state: when we model our own actions in accordance with the desire we have assumed, a desire taken from an impossible other, it is hyperreal—no longer attached to organic reality.

There is a need in advanced capitalism for greater and greater consumption. In order to fill the void there has to be a desire in the consumer population to consume with ever-greater vigor. Baudrillard’s conceptual framework provides us with a powerful tool for understanding and critiquing the logic and influence of advanced capitalism, where Lacan’s psychoanalytic conceptions help to understand why and how the simulacra shape our desires. Just as the family sitcom and pornography provide models of “reality,” and the desire for enactment, upon which family life and sexual encounters are largely

based, it is a rational jump to imagine the capitalist marketing and media machines artificially creating a desire to copy the model of the typical consumer, in much the same way. As long as we are bombarded with imagery of the happy person who owns the next hot new item, it is easy to see how desire will be generated in others to follow suit.

In a sense, this paper has in a long way around come to the point where we can conclude that to satisfy the profit demands of contemporary capitalism, of the producers, media influence and advertising are important and that media influence and advertising work in getting people to buy products. This seems like something we (likely) could have agreed upon before getting off the ground, but what I hope to have shown is *how it is* that these elements can work in generating desires and with them a demand for products. Beyond that, I hope that my account sheds light on how intimately artificially manipulated desires weave into our (sub)consciousnesses. I hope to have given some insight into questions that ask, for example, why it is that a group of students who each feel that rampant consumerism is wrong still max out their credit cards over Christmas? I want to clarify too that even those individuals responsible for production, advertising, the media, etc., are themselves subject to this hyperreal state. The barrage of ideology-in-imagery is such that really no one escapes it. The influence upon individuals of the hyperreal state is penetrating. The reality in which we exist is to a large degree artificial, but in a sense what I call artificial (as opposed to organic) reality is the actual real in which we act. In other words, there is a sense in which the hyperreality of our late-capitalist culture is a fake, but another sense in which this hyperreality is simply real.

Works Cited

- Baudrillard, Jean. *Simulacra and Simulation*. Ann Arbor: The University of Michigan Press, 1994.
- ."Simulacra and Simulations." *Jean Baudrillard, Selected Writings*, ed. Mark Poster. Stanford University Press, 1988, pp.166-184. <http://www.stanford.edu/dept/HPS/Baudrillard/Baudrillard_Simulacra.html>
- Evans, Dylan. *An Introductory Dictionary of Lacanian Psychoanalysis*. London: Routledge, 1996.
- Lacan, Jacques. *Écrits: The First Complete English Edition*, trans. Bruce Fink. New York: W.W. Norton & Company, 2006.
- The Seminar of Jacques Lacan Book XI: The Four Fundamental Concepts of Psychoanalysis*, ed. Jacques-Alain Miller, trans Alan Sheridan. New York: W.W. Norton & Company, 1998.

On Late-Capitalist Desire

Mann, Douglas. *Understanding Society: A Survey of Modern Social Theory*. Don Mills Ontario: Oxford University Press, 2008, pp. 283-292.

Žižek, Slavoj. *How to Read Lacan*. London: Granta Publications, 2006.

The Quest for the Virtuous Life: Can Lao Tzu's Wu Wei be Considered a Virtue?

CHRIS DIMATTEO

TRINITY COLLEGE IN THE UNIVERSITY OF TORONTO

The purpose of this paper is to explore the nature of wu wei (“non-doing”) as explained in Lao Tzu’s Tao Te Ching, and, using a comparison with the nature of virtue in Aristotle’s Nichomachean Ethics, show how the doctrine of wu wei can be considered a virtue. Due to the inherently unknowable nature of the tao, Lao Tzu does not provide a clear, explicit list of virtues, as Aristotle does. Instead of understanding virtue as dispositions to act, as Aristotle suggests, Lao Tzu proposes a disposition to not act. This is not inaction; he advises us to follow the course of nature, and take the path of least resistance in our actions. This must be seen as a virtue, because the doctrine of not acting is the only way in which the ideal life, government, and world can exist. Various Aristotelian objections are presented and refuted to show that, though Aristotle’s conception of virtue is entirely different from Lao Tzu’s, wu wei can still be regarded as such. The “appeal to nature” logical fallacy is raised: why is the doctrine of wu wei, acting in accordance with the natural way, desirable simply because it is “the natural way”? Since, however, the tao is both the reason for our existence, and the source and origin of all goodness and order in the universe, the only way to achieve goodness and order in our own lives is to follow its teachings. The tao is the guiding force of all that we experience; we must not resist it if we are to flourish. Furthermore, this application of logic is also against the tao itself; we must “exterminate learning” and make a leap of faith to the natural world in order to thrive.

In *Tao Te Ching*, Lao Tzu employs poetry and metaphor to explain how one must live in accordance with the *tao*, “the way.” The concept of *wu wei*, translated as “non-doing,” is central to living the *tao*. Lao Tzu contends that by taking the path of least resistance in our daily activities and following the natural order of things, we shall be most successful. This unique characterization of the ideal life is far removed from those found in the western tradition, particularly the virtuous life described in Aristotle’s *Nicomachean Ethics*. The purpose of this essay is to show that, while *wu wei* may not be a virtue in the Aristotelian sense of the word, it has the same ends as Aristotle’s virtues and can be considered a virtue on the basis of its own merits.

In the *Nicomachean Ethics*, Aristotle presents a very clear definition of how one must live a good life. For Aristotle, cultivating virtue means cultivating a habituated state, or disposition to act, that allows us to wilfully and relatively unthinkingly display appropriate character traits at the appropriate times.¹ The character traits of individual virtues are more specifically defined as “a mean between two vices, one of excess and one of deficiency.”² For example, in situations of giving and taking money, generosity is the virtuous mean, while the vice of excess is wastefulness, and the vice of deficiency is ungenerosity.³ Aristotle explicitly enumerates several virtues of character and virtues of thought.⁴ By following these virtues, we will live a flourishing and ultimately just life. Aristotle presents a monist view of ethics: his, and only his, system of ethics and virtues will allow us to flourish. Though it may be a difficult ethical system to follow, Aristotle explicitly tells us what we must do.

From the very beginning of *Tao Te Ching*, it is evident that Lao Tzu will not provide such a clear enumeration of virtues. We know that living in accordance with the *tao* is integral to becoming virtuous: “In his every moment a man of great virtue follows the way and the way only.”⁵ Though the *tao* “was the beginning of heaven and earth” and “the mother of the myriad creatures,” we unfortunately cannot understand or explain it.⁶ The first chapter of *Tao Te Ching* tells us that “[t]he way that can be

¹ Aristotle, *Nicomachean Ethics*, 1106 a 10-15.

² Aristotle, 1107 a.

³ Aristotle, 1107 b 10.

⁴ See *Nicomachean Ethics*, Books III, IV and VI.

⁵ Lao Tzu, *Tao Te Ching*, chapter XXI, line 48.

⁶ Lao Tzu, chapter I, line 1-3a.

spoken of is not the constant way.”⁷ Aristotle’s method of listing off desirable virtues is therefore, with respect to the *tao*, impossible. This would be an attempt at categorizing the uncategorizable or understanding the incomprehensible; any *list* of virtues cannot be in accordance with the *tao*. Instead of a concrete, unchanging set of virtues that we can always follow, we must adopt a natural, submissive set of virtues. This concept is explained further in *Tao Te Ching*: “A man of the highest virtue does not keep to virtue and that is why he has virtue. A man of the lowest virtue never strays from virtue and that is why he is without virtue.”⁸ According to Taoist thought, the truly virtuous man will *not* have a defined set of virtues that he constantly follows; he does not keep to one virtue. Conversely, the man of lowest virtue lists a definite set of virtues in his mind and consistently follows them. By doing this, he is trying to name the unnameable or understand the incomprehensible *tao*. He is not following the true way, and consequently has no real virtue. The Aristotelian concept of virtue is therefore entirely incompatible with that of Lao Tzu. Instead of following a set of instructions, we must attempt to follow the natural *tao*.

Wu wei is a virtue because, according to Lao Tzu, it is the only method through which we can live good lives. Lao Tzu’s *wu wei* has the same goal as Aristotle’s virtues, that is, living a good life. Instead, however, of cultivating a disposition to act, as Aristotle demands, Lao Tzu suggests a disposition to *not* act. Rather than protest or fight nature, we are taught to submit to it, for in this we gain true power: “The submissive and weak will overcome the hard and strong.”⁹ If we are to follow the way in our lives, we must not stand firm against obstacles and attempt to overcome them, for “[t]urning back is how the way moves; weakness is the means the way employs.”¹⁰ In fact, our being submissive will actually allow us to overcome those who are apparently stronger, for “The most submissive thing in the world can ride roughshod over the hardest thing in the world—that which is without substance entering that which has no crevices.”¹¹ Following *wu wei* simply means to follow what is naturally so. Lao Tzu’s disposition to not act, however, is very different from *inaction*. He does not suggest that we do nothing in response to things that happen to us in our lives, but instead, we must

⁷ Lao Tzu, chapter I, line 1.

⁸ Lao Tzu, chapter XXXVIII, line 82.

⁹ Lao Tzu, chapter XXXVI, line 79.

¹⁰ Lao Tzu, chapter XL, line 88.

¹¹ Lao Tzu, chapter XLIII, line 98.

take the paths of least resistance in our actions to be successful. In Chapter LXVIII, he writes, “One who excels as a warrior does not appear formidable; One who excels in fighting is never roused in anger ... This is known as the virtue of non-contention; This is known as making use of the efforts of others...”¹² The act of fighting or being a warrior is not eliminated in *wu wei*. We must not go out of our way to look for a fight, but neither must we neglect to fight when it is necessary. When we face conflicts, we must take the more submissive path to achieve true victory; to be preserved, we must bow down; to be straight, we must bend.¹³

Wu wei can therefore be considered a virtue because it aims to accomplish the same ends as the Aristotelian virtues through different means. Submitting to nature and cultivating a disposition to not act is, for Lao Tzu, the only way we can live a good, flourishing life. We are informed, “[t]hat which goes against the way will come to an early end,”¹⁴ but that “[i]n his every moment a man of great virtue follows the way and the way only.”¹⁵ For one who acts from knowledge of the constant, “to the end of one’s days one will meet with no danger.”¹⁶ Aristotle similarly suggests that the only way we can achieve *eudaimonia* (a term crudely translated as “happiness,” but is something more all-encompassing) is to act in accordance with the virtues he lists for us.¹⁷

The ultimate goal of *Tao Te Ching* is to produce a virtuous ruler, and ultimately, a virtuous society through the methods of *wu wei*. By cultivating the way in the empire, “its virtue will be pervasive.”¹⁸ Lao Tzu observes that people are difficult to govern because authorities are too fond of *action*.¹⁹ Rulers must therefore adopt *wu wei*: if they govern the empire by not being meddlesome, the people will be transformed, rectified, and prosper of themselves.²⁰ Governing an empire is like boiling a small fish, since the simple act of handling it can spoil it.²¹ If citizens of the state adopt *wu wei*, they will find contentment in the way

¹² Lao Tzu, chapter LXVIII, line 166-67.

¹³ Lao Tzu, chapter XXII, line 50.

¹⁴ Lao Tzu, chapter XXX, line 70.

¹⁵ Lao Tzu, chapter XXI, line 48.

¹⁶ Lao Tzu, chapter XVI, line 37.

¹⁷ Aristotle, 1102 a 5.

¹⁸ Lao Tzu, chapter LIV, line 123.

¹⁹ Lao Tzu, chapter LXXV, line 181.

²⁰ Lao Tzu, chapter LVII, line 133.

²¹ Lao Tzu, chapter LX, line 138.

they live their lives, and “the people of one state will grow old and die without having any dealings with those of another”—a peaceful society, even *world*, will result.²² For Lao Tzu, *wu wei* is not only the method through which one forms the ideal individual, but also the ideal ruler, the ideal state, and the ideal world. States will not interfere with each other, rulers will not interfere with their citizens, citizens themselves will not interfere with each other, and everybody will follow the natural order of things. We can quite clearly perceive *wu wei* as a virtue because, for Lao Tzu, the disposition not to act will produce an ideal world.

Some important objections can be made about the conception of *wu wei* as a virtue, and must be dealt with here. According to Aristotle, the qualifications for actions in accordance with virtue are: “[f]irst, the agent...must know [that he is doing virtuous actions]; second, he must decide on them, and decide on them for themselves; and, third, he must also do them from a firm and unchanging state.”²³ This presents two problems for considering *wu wei* as virtue, which will now be dispelled separately.

Firstly, if following the *tao* is the only way of cultivating virtue in our lives, but the *tao* is ultimately unknowable and incomprehensible, how can one ever be in a state that one *knows* one is doing virtuous actions? Though the *tao* may be fundamentally unknowable, Lao Tzu presents us with a number of metaphors and images to show how one can act in accordance with it. The most explicit example of this is in Chapter LV of the *Tao Te Ching*: “One who possesses virtue in abundance is comparable to a newborn babe: Poisonous insects will not sting it; Ferocious animals will not pounce on it; Predatory birds will not swoop down on it; Its bones are weak and its sinews supple yet its hold is firm.”²⁴ We are urged to “Exhibit the unadorned and embrace the uncarved block, Have little thought of self and as few desires as possible.”²⁵ Lao Tzu presents us with an immensely difficult task when showing us how to live in accordance with the way: we are to look at the world as if we were babies again. Due to their lack of experiences, babies have no unnecessary desires, and therefore do not try to alter or fight against the natural order of the world to realize their desires. We are to experience the world with child-like wonder: as a child will unthinkingly follow and submit to his parents, we must follow and

²² Lao Tzu, chapter LXX, line 193.

²³ Aristotle, 1105 *a* 30.

²⁴ Lao Tzu, chapter LV, line 125.

²⁵ Lao Tzu, chapter XIX, line 43.

submit to the natural way of the world. Excessive desire will only force us to attempt to control the world to achieve our desires, which is fundamentally against the *tao*. It is therefore not fair to say that Lao Tzu provides us with absolutely *no* guidance in how to follow the *tao*. Though the virtues of the *tao* may be unknowable insofar as they cannot be explicitly listed and explained, as Aristotle's are, we can understand them through metaphor. This conception of virtue is considerably more difficult to follow than Aristotle's; Lao Tzu seems to acknowledge the inherent difficulty of *wu wei* in Chapter LXX: "no one in the world can understand them [my words] or put them into practice."²⁶ Though we cannot follow the *tao* in the same way as we can follow a list of instructions, we are given metaphorical advice on how to do so. We may not be able to "know" that we are committing virtuous actions, as in the Aristotelian sense, but this is inherent in the fundamentally unknowable *tao*. Therefore, *wu wei* is still a virtue despite the fact that we may have no knowledge that we are committing virtuous actions.

A second objection: if following the *tao* means acting in accordance with nature and not resisting, how could one commit a virtuous action from a "firm and unchanging state"? Simply put, the idea of a firm and unchanging state is entirely antithetical to the conception of *wu wei*. A newborn babe, the exemplar to whom we look for guidance, does nothing from a firm and unchanging state; it does not have the desires necessary to *have* a firm and unchanging state. Being firm and unchanging would imply resistance towards nature, which is, of course, against the *tao*. Both of these objections, derived from Aristotle's writing on virtue, show how *wu wei* may be incompatible with the Aristotelian sense of virtue, but can be considered a virtue on the basis of its own merits.

A further philosophical issue, independent from Aristotle, presents itself when examining *wu wei* as a virtue. *Wu wei*, and Taoism itself, are supposedly desirable because they are *natural*: "Man models himself on earth, Earth on heaven, Heaven on the way, And the way on that which is naturally so."²⁷ Inherent to Taoism, then, is the logical fallacy of "appeal to nature": just because something is natural, it does not necessarily follow that it is desirable or good; conversely, just because something is not natural, it does not follow that it is undesirable, or wrong. If our virtue of *wu wei* is based upon "swimming downstream," for example, just because it is the "natural" way, what if

²⁶ Lao Tzu, chapter LXX, line 170.

²⁷ Lao Tzu, chapter XXV, line 58.

the natural way is not the best possible way for us to live? This idea certainly calls all of *wu wei*'s merits as a virtue into question. For Lao Tzu, though, since the way is responsible for Heaven, Earth, the myriad creatures, and, inevitably, human beings, it only makes sense that we should base our lives around it. The way is the reason we exist (though Lao Tzu explains the way is not strictly causal), but it still forms a fundamental part of us: "The way is to the world as the River and the Sea are to rivulets and streams."²⁸ A rivulet shares some of the characteristics of a river, but it is ultimately much smaller and less important than the river; the world similarly shares some characteristics of the way, but there is *more* to the way than what we experience in the world. Though we can never fully understand the way's power, it is still the reason for our existence, and as such, we must respect and attempt to emulate it. If the way is the sole reason for order and greatness in the world,²⁹ the only way to cultivate order or greatness *in our lives* is to follow the way to the best of our meagre human abilities. Furthermore, it is even arguable that approaching the virtue of *wu wei* rationally, and criticizing its merits as a virtue on the basis of a logical argument, the "appeal to nature," is fundamentally against the way itself. "Exterminate learning and there will no longer be worries,"³⁰ Lao Tzu tells us. We must put the rational paradigm that has been so ensconced in our world aside, make a "leap of faith," and submit to nature. Only then will we adopt the true virtues of *wu wei* and, to the end of our days, we will meet with no danger.³¹

Wu wei must therefore be considered a virtue. Though *wu wei* is an entirely different conception of virtue from Aristotle's, its goal is to live a good life, through acquiring a disposition not to act. For Lao Tzu, the only way to achieve desirable individuals, societies, governments, rulers—and, at the end of the *Tao Te Ching*, there are even hints towards international relations—is through the adoption of the virtue of *wu wei*. Aristotle had the same goals in mind when enumerating his set of virtues in *Nichomachean Ethics*. The two are completely different systems that attempt to achieve the same noble goal through different means, namely, the virtuous life.

²⁸ Lao Tzu, chapter XXXII, line 73.

²⁹ See chapters III, XVI, XXI, XXXIV, LIV, and LIX of *Tao Te Ching* for examples.

³⁰ Lao Tzu, chapter XX, line 44.

³¹ Lao Tzu, chapter XVI, line 37.

Works Cited

- Aristotle, *Nicomachean Ethics*. Trans. Terence Irwin. Indianapolis: Hackett Publishing, 1999.
- Lao Tzu, *Tao Te Ching*. Trans. D.C. Lau. England: Penguin Books, 1963.

Relevant Possibilities, Intentions, and Moral Responsibility

DANIEL MOOSE

UNIVERSITY OF ARKANSAS

*Questions of free will often lead to a discussion of moral responsibility. It is generally assumed in a Libertarian system (one in which all action is free) that a moral agent is responsible for his actions. A Hard Determinism system (one in which no action is free) presents problems for moral responsibility. Since the agent has no control over his actions, it does not always seem most logical to hold him responsible for those actions. In their book *Responsibility and Control: A Theory of Moral Responsibility*, John Martin Fischer and Mark Ravizza discuss a compatibilist system in which some, but not all, actions are free. In cases in which an agent has guidance control, i.e., he is free to act only with respect to some actions but is not free to act with respect to others, assigning moral responsibility can become difficult. In this paper, I attempt to develop a necessary condition for moral responsibility by revising the Principle of Alternate Possibilities (PAP) to address the counter-examples that are brought against it. In the end, I put forth the Principle of All Considered Possibilities (PACP), an intention-based condition that allows a person to be morally responsible even when lacking alternate possibilities.*

One of the simplest and most intuitive proposed necessary conditions for moral responsibility is the Principle of Alternate Possibilities (PAP), which states that a person is morally responsible for what he has done only if he could have done otherwise. However, this principle is easily defeated by Frankfurt-style counter-examples.¹ John Martin Fischer and Mark Ravizza describe a difference between regulative control, where all alternate possibilities are available to an agent, and guidance control, where an agent is free to act only with respect to some actions but is not free to act with respect to others. In this paper, I attempt to develop a necessary condition for moral responsibility by revising PAP to address the counter-examples that are brought against it. In the end, I put forth the Principle of All Considered Possibilities (PACP), an intention-based condition that allows a person to be morally responsible even when lacking alternate possibilities.

1. Regulative Control vs. Guidance Control²

In their book *Responsibility and Control: A Theory of Moral Responsibility*, Fischer and Ravizza begin the discussion of control by offering simple explanations of both regulative control and guidance control. To do this, they offer the example of Sally driving her car. Assuming that her car is working correctly, if Sally were to desire to turn the car to the right and went through the appropriate motions to do so, the car would turn right. Moreover, if Sally were to desire to turn left rather than right and went through the appropriate motions to do so, the car would turn left.

What is important about this case is that not only could Sally have gone right, but she had the choice and the ability to do otherwise (namely, to turn left). In this case, we would say that Sally had regulative control with respect to steering the car. To modify the same case to show her only to have guidance control and not regulative control, we would assume that the car will turn right regardless of Sally's actions. Sally still sits in the driver's seat and can turn the steering wheel as before, but this time, if she were to try to turn left, her action would fail. However, if she were to go through the appropriate motions to turn the car to the right (at the proper time), the car would turn right. Because

¹ Frankfurt, Harry G.. *The Importance of What We Care About*. New York, NY: Cambridge University Press, 2005. pp. 6-8.

² Fischer, John Martin and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. New York, NY: Cambridge University Press, 2000. pp. 30-31.

she still has the ability to guide the car to the right, albeit through a determined path, she is said to have guidance control with respect to her turning right.

2. Frankfurt-Style Cases

a. The Train Case³

The train case offered by Fischer and Ravizza in Chapter 4 lays out a scenario similar to the following. Ralph awakens to find himself on a train hurtling down the tracks. The train is approaching a fork in the tracks with track A on the left and track B on the right. Ralph has the ability (and we're assuming that he has the knowledge) to guide the train onto either track A or track B. Ralph knows that track A leads to Syracuse. Ralph believes that track B leads to Rochester, but track B, in fact, leads to Syracuse as well. Because Ralph is oblivious to the result of choosing track B, he is presented, in his own mind, with the choice between taking track A to Syracuse or taking track B to Rochester.

b. The Mayor Case⁴

The mayor case is similar in nature to the train case. In this case, Sam is faced with the choice of whether to kill the mayor or to refrain from doing so. Jack wishes the mayor to be dead, and so he fashions a device which he implants into Sam's brain such that if Sam were to, for any reason, fail to carry through with his murderous plot, the device would activate and cause Sam to kill the mayor despite his wishes. Sam knows that if he chooses to kill the mayor, he will do so. Sam believes that if he chooses not to kill the mayor, he will not, when he would, in fact, do so. Because Sam is oblivious to Jack's implanting of the device into his brain, Sam is presented, in his own mind, with the choice between deciding to kill the mayor and doing so or choosing not to kill the mayor and not doing so.

3. Relevant Alternate Possibilities

The cases discussed thus far deal with the presence or lack of alternate possibilities. I offer a distinction between alternate possibilities

³ Fischer and Ravizza, pp. 94-95.

⁴ Fischer and Ravizza, pp. 29-30.

in general and *relevant* alternate possibilities. An alternate possibility must meet two requirements for relevancy:

An alternate possibility is relevant if and only if:

- (1) It is present in a nearby possible world, and
- (2) It is meaningful to the situation in question.

First, relevant (or nearby) alternate possibilities are the alternate possibilities which are present in the nearest possible worlds. For example, in the Mayor case, Sam's relevant alternate possibilities are limited to *killing the mayor* and *not killing the mayor*. Although the case is set up such that Sam's only real possibility is to kill the mayor, it is easy to imagine a nearby possible world in which Jack has not implanted the device, and thus Sam has two possibilities. However, we would not consider *flying (like a bird, not in an airplane) away from where the mayor is* to be an alternate possibility for Sam because a world in which Sam can fly like a bird is not very near to our actual world.

The other requirement for relevance of alternate possibilities—as obvious as it may seem—is that the alternate possibility is relevant to the situation at hand. In the Mayor case, we could imagine that Sam has alternate possibilities—other than to kill the mayor and not to kill the mayor—that do not affect whether or not Sam kills the mayor. For example, Sam has the alternate possibilities to wear a hat and not to wear a hat. Excluding some absurd possibility that hat-wearing has a serious effect on Sam's disposition toward killing the mayor, talk of Sam's wearing a hat or not is unproductive to the current discussion of moral responsibility and is thus not worth considering.

4. Moral Responsibility and PAP

a. A Revision of PAP

The two cases mentioned above—the train case and the mayor case—are both similar enough that moral responsibility can be assessed analogously in the two cases. Neither Sam nor Ralph is denied the freedom to choose. In each case the agent is offered two choices, although either choice ultimately results in the same consequence-universal. Ralph can either choose to go to Syracuse or choose to go to Rochester, but he will end up in Syracuse regardless. Sam can either choose to kill the mayor or choose not to kill the mayor, but he will end up killing the mayor regardless.

Although these cases are intended as counterexamples to PAP, I don't think the principle should be thrown out altogether.

In fact, a simple revision to PAP may once again allow it to be used as a system for determining moral responsibility. As given by Harry Frankfurt, the Principle of Alternate Possibilities states:

PAP: A person is morally responsible for what he has done only if he could have done otherwise.⁵

If PAP were to hold, Sam could not be held morally responsible for killing the mayor, regardless of which path he took. Intuitively this seems wrong. How, if Sam chooses to kill the mayor (and thus kills the mayor), is Sam not morally responsible for killing the mayor?

For this reason, I propose my own revision to Frankfurt's PAP, for which I assume a causal relation between an agent's choice and his action.

PAP': A person is morally responsible for what he chooses to do only if he could have *chosen* to do otherwise.

b. Problems for PAP'

This revision of PAP has some serious problems though. Frankfurt cases can easily be written such that no element of choice exists. Just as the two cases above eliminate the agents' freedom to act, it would be just as easy to take away the agents' freedom to choose.

In the mayor case as I presented it, Sam was not free to act other than killing the mayor; however, he still had the freedom to choose whether or not he would do it (although a decision for the negative would have been futile). But what if we went one step further and took away Sam's ability to choose? Imagine now that the same conditions—Sam plans to kill the mayor, has a device implanted that will ensure that outcome—apply, but instead of kicking in after Sam has already chosen, the device would initiate before the point of Sam's choosing. Just as before, the device would not activate if Sam were to choose to kill the mayor. Now, if Sam were to begin to choose not to kill the mayor, the device would activate and cause him to choose to kill the mayor.

In this type of Frankfurt case, the agent is not given any choice to do otherwise. Thus, PAP', as it is worded, is shown to be inadequate for addressing moral responsibility. However, I do

⁵ Harry Frankfurt, p 1.

not take this to be a serious problem for my project. Regardless, the wording of PAP' still needs some further manipulation, so I offer a second revision.

c. Another Revision of PAP

One factor common to the counterexamples is that there are external conditions that limit the agent's freedom. In the train case, the train tracks are already laid down; in the mayor case, the device has been implanted. Without these conditions, moral responsibility would be easy to interpret for the agent. This logic is the basis for yet another revision of PAP:

Principle of Perceived Possibilities (PPP): A person, S, is morally responsible for what he has done only if, given the relevant alternate possibilities perceived by S, S could have done or could have chosen to do otherwise.

In my explanations of the cases before, I was careful to point out that each agent perceived there to be two options from which he had to choose. The two previous versions of PAP failed to address this illusion of choice presented to the agent, and because of this, they were defeated. By viewing an agent's actions relative to a consequence universal, we unfairly attribute the external to an agent, ignoring the internal.

Not only is this unfair to the agent, it seems to be a highly inaccurate method for assigning responsibility. Moral responsibility should be assigned based on the qualities, actions, choices, et cetera of the agent, not on situations over which the agent has no control. For instances in which Sam kills the mayor on account of Jack's device, the action or choice is obviously not Sam's; therefore, Sam should not be held responsible in these cases.

PPP, however, moves from the unalterable circumstances to the agent's own perception of his situation. If an agent perceives there to be two options for a given situation, he should be held morally responsible only for choices or actions which could be applicable to him regarding those options, even if some of those choices or actions could not come to pass due to certain circumstances (such as Jack's device). Another way to think of this is to say that if the unperceived circumstances were not to apply, an agent could be held morally responsible for any choices or actions he might take.

By viewing the agent's situation like this, we remove all consequence universals from the equation. In the train case,

Ralph is no longer just responsible for choosing track A or track B; he is responsible for his decision between Syracuse and Rochester. In the mayor case, Sam is now responsible for choosing between killing the mayor and not killing the mayor. Because, when we factor in the unperceived circumstances, Sam does not actually have the freedom to choose (based on the second version of the mayor case), we must rely on his disposition or his intention with regards to killing the mayor. If Sam intended to kill the mayor, in all cases, he is morally responsible for killing the mayor. However, if Sam's unaltered intentions were not to kill the mayor—although he does so regardless—he is not morally responsible for killing the mayor.

5. Expanding PPP

a. Shortcomings of PPP

As shown above, PPP covers the Frankfurt-style cases, thus allowing a person to be morally responsible for what he has done even in the absence of alternate possibilities. So long as the person perceives an alternate possibility for himself, he is responsible for his actions in respect to the alternate courses of action he believes to exist. Unfortunately, counterexamples exist to test PPP.

The Maze Case

Imagine the scenario in which a guy, Steve, is put in a maze. There are two possible paths he could take. Path 1 leads to an apple; Path 2 leads to an orange. Steve knows to what each path leads. Moreover, Steve knows that the person who put him in the maze has blocked off Path 2 such that the orange is unattainable to Steve. Further assume that Steve cannot just choose to stay where he is (and he is aware of this fact). Steve has only one path which he can choose, Path 1, which leads to the apple. Because Steve is aware, in his own mind, of the complete scenario, he is not presented with choice as to what he will do.

b. Parrying the PPP

Initially, the Maze Case seems a very straightforward situation, in which Steve cannot be held morally responsible. Steve does not have any alternate possibilities. Furthermore,

Steve does not even perceive himself to have alternate possibilities. Thus he can be held morally responsible by neither PAP nor PPP.

However, there are possible arrangements of this scenario in which, I would argue, Steve should be held morally responsible for proceeding down Path 1. Imagine that Steve is reflecting on his situation. He realizes that his only option is to take Path 1 to the apple. However, he also realizes that the other options might have existed (and, in fact, do in another possible world). Upon reflection, Steve decides that he would, no matter what other options were available to him, choose Path 1 to the apple.

In this case, Steve's intentions line up with his only available option. Intuition seems to dictate that we would want to hold Steve morally responsible for choosing the apple if choosing the apple were what he truly wanted to do, regardless of whether or not alternate possibilities existed; however, most views of moral responsibility, including PPP, would say that he is not morally responsible in this situation because he neither has nor perceives alternate possibilities for himself. Since Steve knows that he doesn't have any alternate possibilities, he doesn't fall under the "given the circumstances perceived by S" clause of PPP. The alternate possibilities in this case are not perceived; rather, they are conceived.

Another situation in which a person's intentions align with what he perceives to be his only option is the following scenario.⁶ John goes to bed every night and closes the door to his room. When he wakes up in the morning, he opens his door and is able to leave the room. At some point, John discovers that his roommate Mark, who is somewhat of a creepy guy, has been monitoring his sleep. At the instant when John goes into his room at night to sleep, Mark locks the door such that if John were to try to leave the room, he could not. In the morning, Mark unlocks the door at the very instant John wakes and wishes to leave the room.

For the time period during which Mark has the door locked, John is not free to leave the room. However, John intends to be in the room for that time. John knows that even if

⁶ This scenario is my version of the scenario offered by John Locke in Book II, Chapter 21, Section 8 of *An Essay Concerning Human Understanding*.

Mark were not to lock the door, he would remain in the room. Should he not also be considered responsible for being in the room because it is what he wants?

c. An Alternate Version of PPP

The problem that arises from this new style of counter-example is that the person is aware that he does not have any alternate possibilities, but he still conceives of possibilities that he might have in a nearby possible world. Despite this awareness of his situation, the person should still be held responsible when his intentions align with his actions, even when no possible alternate actions exist. To account for this, I offer an alternate version of PPP.

Principle of Conceived Possibilities (PCP): A person, S, is morally responsible for what he does only if, given the relevant alternate possibilities conceived by S, S would still do that which he does.

Just as when I argued for PPP, when I was careful to point out that Sam and Ralph perceived themselves to have two possibilities, I have been careful to show that Steve and John conceive of alternate possibilities for themselves. Steve, knowing he must proceed down Path 1, imagines a similar situation in which he could take Path 2; John, knowing he must stay in the room, imagines a similar situation in which he may come and go as he pleases. However, if Steve realizes this and still knows he would choose Path 1 in any situation (and likewise John would stay in the room in all cases), then he should be morally responsible for choosing Path 1. PCP allows us to hold Steve responsible.

6. Necessary Conditions for Moral Responsibility

For cases in which a person truly has alternate possibilities, or to put it in Fischer and Ravizza's terms, has regulative control, it is simple to say that he is morally responsible for that which he does. However, there are many realistic cases in which a person does not have alternate possibilities, yet we would want to hold him morally responsible for his choices and actions. In these cases in which the person has only guidance control, he can still be held responsible even when he has no other option but to act as he does. The cases above have illustrated three possible conditions for candidacy for moral responsibility.

A person is morally responsible only if:

- (1) He (correctly) perceives himself to have relevant alternate possibilities; or
- (2) He (incorrectly) perceives himself to have relevant alternate possibilities; or
- (3) He perceives himself not to have relevant alternate possibilities, but, if he were to perceive that he has relevant alternate possibilities, he would act no differently.

Condition (1) can be illustrated by the case of regulative control in which Sally has the freedom to turn the car left or right. Condition (2) can be illustrated by the guidance control cases of Sam and Ralph, where they perceive themselves to have alternate possibilities when they, in fact, have none. Condition (3) can be illustrated by the Maze case and the Locked Room case, where the agent perceives no alternate possibilities but, in conceiving of them, realizes he would act no differently than he does.

Having established the types of scenarios in which we might assign moral responsibility, we must now turn to the principles by which we determine whether or not an agent is morally responsible. Condition (1) is the most simple. If a person fulfills (1), he is determined to be morally responsible by PAP. Condition (2) is covered by PPP. PPP allows that, even if the agent does not have alternate possibilities, he can be held responsible for the alternatives he perceives (even though his perceptions are false). In condition (3) the agent is held morally responsible despite his recognition of the fact that he has no alternate possibilities. PCP, though, exposes that the action in question is consistent with the agent's intentions and thus the agent should be morally responsible for that action.

For now, we have three conditions and three principles for moral responsibility. This seems a bit cumbersome. To have an effective method for determining moral responsibility, it would be much nicer if the three principles could be combined into one principle. Fortunately, PPP is a revision of PAP, and, in revising it, I was careful to incorporate the original principle into the revision. PPP covers condition (2) for the reasons argued in the sections above; it covers condition (1) because of how it is worded. The phrase, "given the circumstances perceived by S," which was primarily intended to explain how a person could be responsible when his perceptions are wrong, also works for cases in which the person's perceptions are correct. In cases of (1), the "circumstances perceived by S" are the correct circumstances, which include alternate possibilities for S.

PCP is more difficult to incorporate with the other two theories. Rather than being a further revision of PAP (as PPP is), it is an altered version of PPP which replaces the notion of perception with the notion of conception. PPP by itself accounts for cases of the nature of (1) and (2) but does not account for cases of the nature of (3); PCP by itself accounts for cases of the nature of (3) but does not account for cases the nature of (1) and (2), because its description of the conceptions implies that what the agent conceives is contrary to that which he actually perceives. Therefore, a good marriage of PPP and PCP should yield a principle that can serve as a tool for assessing moral responsibility in all of the above cases. In an attempt at that, I offer the following principle.

Principle of All Considered Possibilities (PACP): A person, S, is morally responsible for what he does only if, given the relevant alternate possibilities conceived or perceived by S, S would still do that which he does.

PACP, just like PPP, makes no judgment concerning the validity of the alternate possibilities perceived by the agent. For this reason, it is still fit to assess situations of types (1) and (2). The phrase “given the alternate possibilities conceived” incorporates type (3) situations. PACP is simply a combination of the important parts of PPP and PCP.

The notion of relevant alternate possibilities is also preserved in this latest revision. To take the Maze case, Steve could conceive of an alternate possibility whereby he could simply fly out of the top of the maze. However, Steve does not have the ability to fly, so this alternate possibility is a bit too much of a stretch to be helpful for the goals of PACP. Only the alternate possibilities available in the nearest possible worlds should be considered in assessing moral responsibility, that is to say that an alternate possibility is only relevant if it occurs in nearby possible worlds. In this way, moral responsibility is based on the rational intentions of the agent in question.

Defining Goodness

SIMONA AIMAR

UNIVERSITY OF ST. ANDREWS

*G. E. Moore's Open Question Argument has it that there is no satisfactory definition of the term "good." Cornell realists purport to meet Moore's challenge by adopting an externalist semantics of ethical terms. On the face of it, non-cognitivist point out that an externalist theory of meaning cannot account for the practical function of moral discourse. This paper suggests (1) that there are two different cases for the indefinability of goodness: a Moorean version and a Humean (or non-cognitivist) version; and (2) that an apt modification of Cornell realism, as the one offered by Mark Van Roojen, can meet both versions of the Open Question Argument.**

* Many thanks to Sarah Broadie and Julien Murzi.

How is the term good to be defined? This question was addressed by G.E. Moore in his *Principia Ethica*.¹ By means of the so-called Open Question Argument, Moore claimed that no naturalistic definition can account for ethical terms. His positive thesis was that goodness is a *sui generis* property, accessed by intuition.² Moore's Platonic ontology and its mysterious epistemology (what is "intuition" after all?) did not convince philosophers such as A.J. Ayer³ and R.M. Hare.⁴ In their view, moral properties are indefinable because they stand for no properties at all. Terms such as "good," they argued, essentially have a *practical*, rather than a descriptive, function. The metaethical position they ended up with—non-cognitivism—accounted for ethical language by considering moral judgements as an expression of our moral attitudes. So-called Cornell realists,⁵ appealing to Kripke and Putnam's new wave semantics⁶ for natural kind terms, claimed that "good" may be identified *a posteriori* with natural properties, thus embracing the view that there are real moral properties. They overcame Moore's challenge for the indefinability of moral terms by adopting an externalist semantics. The success of Cornell realism has been put in question by authors such as, among others, Simon Blackburn.⁷ Roughly, the charge is that Cornell realism is essentially unable to account for the practical function of moral terms. Hence, non-cognitivism should be preferred.

The intention of this paper is that of supporting two theses. The first is that there are two cases for the indefinability of goodness that can be illustrated as corresponding to two versions of the Open Question argument: a Moorean version and a Humean (or non-cognitivist) version. The second thesis is that the Cornell Realist's semantic treatment of moral terms may be suitably amended so to become immune to both versions of the Open Question Argument. The structure of the paper is as follows. In the first section, I trace the Moorean and the Humean versions of the Open Question Argument. The second section considers

¹ Moore: 1903, ch.1.

² Ibid.

³ Cf. Ayer: 1936, ch. 6.

⁴ Cf. Hare: 1952, ch. 5.

⁵ The label "Cornell realists" is generally applied to the following authors from New York State: Nicholas Sturgeon, Richard Boyd and David Brink. Their metaethical views are mainly in Sturgeon: 1984, Boyd: 1988 and Brink: 1989.

⁶ Cf. Kripke: 1971 and Putnam: 1975.

⁷ Cf. Blackburn: 1993 and 1998, pp. 119-121. A charge to Cornell Realism also comes from Terence Horgan and Mark Timmons (see footnote 22).

Cornell realism. The third section focuses on a recent proposal by Mark Van Roojen.⁸ It will be suggested that Van Roojen's account of moral terms passes both the open question tests, insofar as it revises Cornell realism by adopting a non merely externalist semantics of ethical language.

1. Non Defining Goodness: the Two Versions of the Open Question Argument

1.1 The Moorean Version: "Good" Is Unanalysable

Moore's Open Question Argument (henceforth, OQA)⁹ purports to show that goodness cannot be aptly defined. In order to show the inadequacy of a definition of "good," Moore suggests the following procedure. Let D be any supposed reductive definition of any moral predicate M . Then, for any competent speaker S and for any object x falling under D , it is always possible that S knows that x is D and yet wonders whether x is also M . In other words, the fact that x is D does not appear to settle, in principle, the question whether x is M . A competent speaker, Moore contends, can always pose the question "is A good?" in front of *any* reductive description of an object A , without betraying conceptual confusion. Hence, it seems to follow, any such reductive definition of M is bound to be inadequate. On the assumption that "good" has a meaning and that a suitable definition is an analysis that exhaustively accounts for the meaning of a term, Moore concludes that no definition of "good" can be given.

Another way of putting the argument may be the following. Moore assumes that two identical properties render identity statements equating them tautological. Accordingly, he has it that a sentence of the form "good = x " should appear to us as a tautology when the variable x is substituted with the property-term equivalent with goodness. But no property, Moore argues, can do this job. Saying, for instance, that "good is pleasure" does not represent a tautology, because we can intuitively recognize that asking ourselves whether pleasure is good does not correspond to doubting whether pleasantness is pleasantness. Assuming that "good" stands for a property and that identity statements rely on conceptual connections, then, Moore invites us to the conclusion that no property is identical with goodness.

As it stands, the argument has raised a number of worries. One of

⁸ Van Roojen: 2005.

⁹ Moore: 1903, ch. 1 (especially §12-13).

the most serious charges is probably due to W. K. Frankena.¹⁰ He argued that the OQA begs the question, since it already assumes what it intends to show, namely that the question stays open. This charge may be met, however, by considering the OQA as relying on our intuitions about the answer to the question “Is x good?” Thus understood, the OQA becomes a challenge—rather than a conclusive argument—for any definitional attempt of goodness.

1.2 The Humean Version: Ethical Terms as Practical

It seems to me that a specific path of interpretation of the Open Question Argument started with Ayer’s *Language, Truth and Logic*,¹¹ and it now represents a common hallmark of the non-cognitivist approaches to metaethics. Non-cognitivists hold that moral judgements are essentially linked with the practical attitudes (as opposed to cognitive/representational states of mind) of the speakers. One of the ways in which they support this view is that of endorsing an interpretation of Moore’s Open Question Argument that I shall dub the “Humean version of the Open Question argument” (henceforth, HOQA).¹²

The authors endorsing the HOQA consider Moore’s argumentative tool as a way of indicating the peculiar function of ethical language. The reason underpinning the non-definability of moral terms

¹⁰ Frankena: 1939.

¹¹ Cf. Ayer: 1946, especially pp.103-106, 108; see also the introduction of the same work, where Ayer claims that ethical terms have a different kind of meaning (ibid. p.15). It might be argued that Ayer’s theory of meaning only depends on his endorsement of the verification principle of Logical Positivism, according to which a sentence is truth-apt only if it is either analytically true or empirically verifiable. However, in respect of ethical terms Ayer joins, in several passages, such positivistic theory of meaning with considerations about the peculiar function of ethical language, thus arriving to read the Open Question Argument in the way in which I would like to emphasize here. Nevertheless, it must be admitted that Ayer seems to waver between being an eliminativist about moral discourse and remarking its distinctive function. The reading of the Open Question Argument I shall consider here is very explicitly expounded in subsequent authors as Hare (Hare: 19 ch. 5) and Blackburn (Blackburn: 1998, p. 70). There is also a case of non exclusively non-cognitivist authors who read Moore’s OQA in this way (see Darwall, Gibbard and Railton: 1992, p. 4).

¹² The label comes from Jackson (Jackson: 1998, pp. 151-153). As far as I know, Jackson is the only author that explicitly talks of two versions of the Open Question Argument. My interpretation of the HOQA might differ slightly from Jackson’s.

as “good,” they contend, relies in the fact that ethical judgements play a commendatory role. In their view, the question “Is x good?” stays open – no matter what descriptive account of x we are given – because that is *not* a question about factual descriptions. Asking whether something is good corresponds to asking whether the object of our judgement is something that should be brought about. The Moorean doubt about goodness concerns the commitment of the speaker towards commending (or condemning) a certain behaviour. Hence, it does not come as a surprise that no reductive description can settle the question whether a certain action, or a certain object, enjoys a given moral feature M .

If the HOQA is considered as correct, it represents a challenge for metaethical theories besides the original Moorean version. So understood, the case for the indefinability of goodness becomes also a case for the practical role of ethical terms. The openness of the question is here implanted within a distinction between statements of facts and statements of values, according to which the point made by the open question argument is that of putting an emphasis on the normative function of moral terms.

2. Towards a Post-Moorean Definition of Goodness: Cornell Realism

2.1 Trying to Overcome the Open Question Argument by Means of New Wave Semantics

I shall here focus the attention on the account for ethical language proposed by Richard Boyd,¹³ David Brink¹⁴ and Nicholas Sturgeon.¹⁵ Their view, usually referred to as Cornell realism (henceforth, CR), accounts for the meaning of ethical terms by endorsing a brand of ethical naturalism. According to CR, there are moral properties and these are nothing but a sort of natural facts. Consequently, moral judgements are perfectly truth-apt, since our usage of moral terms is regulated by referents concretely existing out in the world. So described, CR appears to be an easy target for Moore’s OQA. Its suggested reduction of moral properties to natural facts seems to entail precisely the descriptivist position about ethical terms to which Moore objected. On the face of it, Cornell realists contend it is possible to countenance a post-Moorean version of naturalism and, at the same time, accept that the conclusion of Moore’s argument. The crucial motivation

¹³ Boyd: 1988.

¹⁴ Brink: 1989.

¹⁵ Sturgeon: 1984.

for their claim relies in relatively recent developments within the philosophy of language, firstly suggested by Saul Kripke¹⁶ and Hilary Putnam.¹⁷

To put it roughly, the idea is that moral terms rigidly refer to natural properties—just as, in Kripke and Putnam’s view, “gold” and “water” do.¹⁸ Reference is causally secured: the reference of moral terms is determined by the causal connections relating the speakers’ usage of a term and its referent. Once the reference has been fixed, a causal interaction between the linguistic community and the item in question determines the referential property of the term itself. As Richard Boyd puts it:

Roughly, and for non-degenerate cases, a term *t* refers to a kind (property, relation, etc.) *k* just in case there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term *t* will be approximately true of *k* (excuse the blurring of the use-mention distinction). (Boyd: 1988, p. 116)

According to Boyd, a cluster of natural properties¹⁹ and a mechanism of causal interaction with socially coordinated epistemic accesses underlies our usage of a moral term *t*. Moral properties, thus, are accounted as those natural properties that have an explanatory (causal) role for our moral talk. Which properties and mechanisms have to be indicated in order to define *t*, Cornell realists contend, is an a posteriori issue.

2.2 The Test of the Moorean Version of the Open Question Argument

If we grant the extension of Kripke and Putnam’s semantics for natural kinds to moral terms, then a (natural) response to Moore’s Open Question Argument might be at hand. The idea is simple. The referents of natural kind terms are only a posteriori knowable. Hence, lack of knowledge thereof has nothing to do with competent speaker’s knowledge. After all, there were competent speakers even before it was

¹⁶ Kripke: 1972.

¹⁷ Putnam: 1975.

¹⁸ Boyd is the author that most develops the semantic thesis of CR in detail. In the next paragraphs I shall follow his version of CR (cf. Boyd, 1988, pp. 114-128).

¹⁹ Specifically, Boyd talks of “homeostatic properties”: cluster of properties contingently grouped, that *homeostatically* tend to occur together. As natural kind properties, they are empirically discoverable and causally efficacious (cf. Boyd: 1988, p 116-117).

actually discovered that “Water = H₂O.” As a result, CR appears to be able to face without fear the OQA.²⁰ Since we become acquainted with the meaning of terms such as “good” *a posteriori*, no conceptual linkage indicates us the possibility of its identification with natural properties. Hence, the openness of the Moorean question “Is *x* good?” may be due to our ignorance about which properties causally regulate our employment of moral terms. A perfectly competent speaker may use a term correctly without needing to be aware of which properties can be synthetically equated with the term in question, and thereby legitimately put in doubt the identification of the two.

2.3 The Test of the Humean Version of the Open Question Argument

One of the difficulties of the theory of meaning put forward by CR emerges when the view is appraised in the light of the HOQA. Non-cognitivists criticize the Cornell realists’ suggested synthetic identifications as insufficient to account for the *normative import* of ethical terms.²¹ The idea is that there is an aspect of the meaning of moral language which is essentially not captured by the causal theory of reference. According to a non-cognitivist, in front of any naturalistic description—although *a posteriori* given—of goodness, the question “Is *x* good?” stays open, because that is a *normative* question about the attitude the speaker should take towards *x*, and *not* merely an inquiry about factual properties.²²

²⁰ Cf. Boyd: 1988, p. 119.

²¹ Cf. Blackburn: 1998, pp. 119-121.

²² This point seems to be made also by Terence Horgan and Mark Timmons’ “Moral Twin Earth Argument” against CR. They attack Boyd’s account for extending Kripke and Putnam’s new wave semantics to moral terms by means of a “revised version of the Open Question Argument.” Their argument asks us to imagine two moral twin earths identical to each other except in the causal factors which determine the meaning of ethical terms as “good.” On one planet the use of “good” is regulated by consequentialist properties, whereas on the other planet people’s use of “good” is regulated by non-consequentialist properties. It seems that if we imagine a meeting between the inhabitants of the two planets, and we think of them as discussing about whether something is good, we plausibly deem they can have a genuine disagreement. Such an argument may be considered as a version of the HOQA, since the disagreement between the inhabitants of the two worlds seems to be precisely a practical dispute about what we ought to do. Timmons and Horgan explicitly indicate Hare’s version of the HOQA as an ancestor of their argument (cf. Horgan and Timmons: 1991 p. 166 n. 22).

Cornell Realists reply to this worry by regarding the practical role of moral discourse as *non* constitutive. They deny the motivational internalism about judgements that the non-cognitivist interpretation of HOQA seems to entail.²³ Moral judgements, they argue, do not have to be interpreted as *internal*, that is, conceptually connected with our motivation to act. The motivational view thereby advocated, usually defined as externalistic, has it that the action-guiding function of moral judgements depends on *contingent* psychological factors. We tend to consider ethics as action-guiding because the psychological traits that make it normative are “deeply seated and widely shared.”²⁴ But there is no intrinsically practical role of moral discourse. In addition, Cornell Realists often attack their internalist opponents in the following way. It is possible to conceive, they contend, of an *amoralist* who judges of an action *x* as morally good and yet lacks motivation to act accordingly.²⁵ We can think that there are agents as Plato’s Thrasymachus, who competently engages in moral debates without being moved to pursue moral actions. Given this counterexample, internalism about judgement seems defeasible. For there may be cases whereby a (necessary) linkage between morality and conduct does not hold.

The dialectic that here takes place between non-cognitivists and Cornell realists may be regarded as leaving both parties with half a victory. The amoralist case, if accepted, has the merit of putting some pressure on the HOQA. The conceptual connection between moral appraisals and action-guidance, by means of which the HOQA is generally explained, encounters the plausible counterexample of possibly conceivable agents for whom the action-guiding role of moral discourse does not obtain. On the other hand, the Cornell Realist’s answer to the HOQA seems to disrespect our moral intuitions. We plausibly retain that in front of any factual account of some situation the possibility of wondering about the morality of pursuing it stays open.²⁶

3 Cornell Realism Revisited: Van Roojen’s Proposal

²³ The exponent of CR that most explicitly accounts for this move is Brink: 1989, ch. 3.

²⁴ Brink: 1989, p. 49.

²⁵ Cf. Brink: pp. 46-50.

²⁶ Blackburn interestingly puts forward a second related objection against the externalist semantics of CR: if to argue about morality corresponds to employ terms regulated by natural facts, the objection runs, we should conclude that in a bad world nobody can think or dream of justice (Blackburn: 1993, pp. 203-204).

3.1 A “New New Wave Semantics”

In the present section I shall assume (i) that the Humean version of the Open Question Argument correctly emphasizes an aspect of ethical language and (ii) that such aspect needs to be explained somehow with an internalistic theory of motivation. Accordingly, I shall support a possible revision of CR that has been recently proposed by Mark Van Roojen.²⁷ The suggestion is that of modifying some points of Boyd’s theory of meaning for moral terms. Such a proposal seems to show that a suitable amendment of the semantic thesis of CR may lead to account for the naturalistic aspects of the meaning of moral terms as well as for their intrinsically practical features.²⁸

Van Roojen’s account, which he labels “New new wave theory”,²⁹ aims at preserving the *a posteriori* definability of moral terms in a broadly Kripkean semantic spirit. Van Roojen identifies the referents of moral terms with sets of natural properties.³⁰ We have seen that Boyd understands the meaning of ethical terms as a *mere* function of natural facts, along the lines of a radically externalist semantics. Van Roojen argues that a sort of naturalistic semantic of ethical terms may be explained along the lines of a different process of reference-fixing, so to avoid an utterly externalist referential theory. Specifically, his idea is that of replacing the notion of reference-determining causal facts with that of reference-determining *epistemic facts*. He suggests that we explain the

²⁷ Van Roojen: 2005.

²⁸ Van Roojen puts forward his proposal with the purpose of giving an answer to the so-called “Moral Twin Earth Argument”, given by Horgan and Timmons (cf. Horgan and Timmons: 1991). Since I regard this argument as a version of the HOQA, I will proceed in my exposition by referring to the latter only (see footnote 22). Van Roojen too seems to read the “Moral Twin Earth Argument” as connected with the practicality of moral discourse (cf. p. 175 n 14). It may be worthwhile to notice that I shall particularly emphasize the aspect of Van Roojen’s proposal that answers to the OQA and the HOQA as presented below. This choice of emphasis may perhaps differ from Van Roojen’s specific intentions.

²⁹ Van Roojen: 2005, p.174.

³⁰ Van Roojen also proposes to account for the notion of “natural” as “discipline-relative,” so that the criteria of “naturalness” may not be the same in, say, physics and morality. In this section, I limit my analysis to the consideration of Van Roojen’s proposal in respect of the semantic features that the OQA and the HOQA seems to underline, although Van Roojen’s observations are worthwhile to be mentioned (Van Roojen: 2005, pp. 180-183).

referentiality of moral terms by means of an *epistemic regulation*. Mimicking Boyd, he puts it in the following way:

Roughly, and for non-degenerate cases, a term *m* refers to a property *p* just in case there exist *epistemically relevant procedures* whose tendency is to bring it about over time, that what is predicated of the term *t* will be approximately true of *k* (excuse the blurring of the use-mention distinction). (Van Roojen, 2005, p. 176 – emphasis added)

Van Roojen understands his notion of “epistemically relevant procedures” in terms of relations between speakers and the world that are responsible for the determination of the semantic value of ethical terms. Not every epistemic relation, he argues, has to be shaped in terms of merely causal relations—as in Boyd’s proposal. Epistemic processes may involve *a priori* conceptual linkages: there may be aspects of *a priori* knowledge of the referent of a term *t* that may play an essential role in determining the referential capacity of the term in question. The referential determination process may rely on a regulation which does not only depend on external properties.³¹ Accordingly, the individuation of the referent of a moral term may turn out to be the result of the interplay between *a priori* referential intentions of the speaker and *a posteriori* factors.

3.2 The Open Question Tests

Van Roojen’s proposal seems to pass both versions of the Open Question Argument. For one, it appears to be immune to Moore’s version of the OQA, since it denies that *a priori* descriptions can be sufficient to pick up automatically the referent of moral terms. For another, the *a priori* features of the reference-determining processes introduced by Van Roojen allow him to incorporate in his theory a loose internalist position. Differently from standard versions of CR, Van Roojen’s view can endorse a conceptual connection between accepting a moral judgement and being motivated to do what it recommends. For the individuation of the reference of moral terms depends here not only on the *a posteriori* causal regulation of natural facts, but also on the referential intentions of the speakers. Resting on the assumption that moral terminology is somehow action-guiding, the use of ethical terms can then be explained as entailing some *a priori* knowledge of the referent, such as that of being the property of an action that it makes sense to do.³²

It can be noticed that the practicality of moral judgements here

³¹ Ibid., pp. 176-179.

³² Van Roojen: 2005, p. 189 ff.

endorsed does not have to account for *every* speaker. Some features actually contributing to the meaning of moral terms may be exemplified only by a sub-set of the linguistic community. Amoralists might be able to make moral judgements and yet not being motivated by them, because endowed with wrong referential intentions. They may be considered as speakers with partial linguistic competence of ethical vocabulary, sufficient to enable them to put forward moral judgements. Analogously to Burge's patient who wrongly thinks of having arthritis in his thigh and yet has enough linguistic competence to use the term "arthritis."³³

Conclusion

I have argued that Moore's OQA and its non-cognitivist interpretation, the HOQA, represent two correlated cases for the indefinability of ethical terms. The OQA shows that reductive definitions of goodness do not account for the nature of "good" in an adequate way. The non-cognitivist reading of Moore's argument, the HOQA, has it that ethical terms are not definable because they have an essentially practical function. Given these two cases for the indefinability of goodness, Cornell realists respond to the challenge of the Moorean OQA by pointing to the possibility of a posteriori naturalistic definitions of ethical properties. Such a proposal, however, risks not to guarantee an account for the practical function of ethical terms—if this is taken to be a constitutive feature of moral discourse. I have suggested that a more satisfactory account for the meaning of ethical terms is obtained by modifying CR externalist semantics. If we introduce a priori referential constraints in the reference-determining processes of ethical terms, as suggested by Van Roojen, it seems possible to relate natural properties and moral terms in a non-counterintuitive way. The proposal allows one to account for ethical language as normative, albeit also relating it with a posteriori factors. Interestingly, Van Roojen's appeal to conceptual links as well as to a posteriori features may perhaps also explain why both merely analytic and merely synthetic accounts for the meaning of ethical terms seem unable to do the job. They might be, so to speak, (mistakenly) one-sided.

There is no intention to claim that the analysis of the positions here considered has been exhaustive, nor to maintain that the reasons given in favor of a revision of CR semantics are conclusive. Rather, I hope to have given support to the thesis that a revision of Cornell Realism might represent an interesting route for accounting for the *prima*

³³ Cf. Van Roojen: 2005, p. 190. And Burge: 1979.

facie indefinability of “good.”

Works Cited

- Ayer, A. J. (1946). *Language, Truth and Logic*. The Camelot Press Ltd: London and Southampton, pp. 5-22 and ch.6.
- Blackburn, S. (1993). “Just causes” in his *Essays on Quasi-Realism*. pp. 199-209.
- (1998). *Ruling Passions*. Clarendon Press: Oxford, ch. 3-4.
- Boyd, R. (1988). “How to be a moral realist?” in Darwall, Gibbard and Railton (eds) *Moral Discourse and Practice* (1997). Oxford University Press: Oxford, pp. 105-135.
- Brink, D. (1989). *Moral Realism and the Foundations of Ethics*. New York: Cambridge University Press, ch. 3-4 and 6.
- Darwall, S. Gibbard, A. and Railton, P. (1992). “Toward *Fin de siècle* Ethics: Some Trends” in their *Moral Discourse and Practice* (1997). Oxford University Press, pp. 3-47.
- Frankena, W. K. (1939). “The Naturalistic Fallacy” in *Mind* 48, pp. 464-477.
- Jackson, F. (1998). *From Metaphysics to Ethics*. Clarendon Press: Oxford, ch.6.
- Hare, R. M. (1952). *The Language of Morals*. Clarendon Press: Oxford, ch. 5.
- Horgan, T. and Timmons, M. (1992). “Troubles for New Wave Moral Semantics: the ‘Open Question Argument’ Revived”, *Philosophical Papers*, 21 (1992), pp. 153–75.
- Kripke, S. (1972). *Naming and Necessity*. Harvard University Press: Cambridge, Mass.
- Moore, G.E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.
- Putnam, H. (1975). “The meaning of meaning,” in his *Mind, Language and Reality*. Cambridge University Press: Cambridge, pp. 215-271.
- Sturgeon, N. (1984). “Moral Explanations” in Copp and Zimmerman (eds.) *Morality, Reason and Truth*. Rowman and Littlefield, pp. 337-352.
- Van Roojen, M. (2005). “Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument” in Shafer-Landau (ed.), *Oxford Studies in Metaethics*, pp. 161-193.

SOPHIA XII SUBMISSIONS

Sophia is always looking for submissions. Information on the requirements for submission will be posted at <http://web.uvic.ca/philosophy>. The deadline for submission is typically near the end of January. Submissions are required to be of a philosophical nature, should not exceed 4000 words and must be accompanied by an abstract. Electronic submissions can be sent as an attachment via email to sophiajournal@gmail.com. Paper submissions can be sent to

Sophia 2009

c/o Department of Philosophy - University of Victoria

P.O. Box 3045

Victoria, BC

V8W 3P4