Florida Philosophical Review: The Journal of the Florida Philosophical Association is an anonymously refereed, electronic journal published twice a year by the University of Central Florida Department of Philosophy.

Florida Philosophical Review has its roots in the Florida Philosophical Association, one of the largest and most active regional philosophy associations in the United States. For several years, the Florida Philosophical Association envisioned a scholarly publication that would support the professional interaction of philosophers in Florida, the enhancement of philosophical education in Florida, and the development of philosophy both within and beyond Florida. *Florida Philosophical Review* realizes that vision and is committed to respecting and encouraging diverse philosophical interests and diverse philosophical approaches to issues while demonstrating the value of philosophy in the contemporary world.

Florida Philosophical Review: The Journal of the Florida Philosophical Association (ISSN 1535-3656) is published twice a year (in June and December) by the University of Central Florida Department of Philosophy, which assumes no responsibility for statements expressed by the authors. As an electronic journal, *Florida Philosophical Review* is distributed free of charge on-line. Individuals or institutions desiring a CD version of the journal, may purchase such for $35.00

Please address all business and editorial correspondence to fpr@mail.ucf.edu or to Shelley Park and Nancy Stanlick, Editors, *Florida Philosophical Review,* Department of Philosophy, Colbourn Hall 411, University of Central Florida, Orlando, FL 32816-1352.

Editors' Introduction

This issue of the *Florida Philosophical Review* includes articles that cover many different areas of philosophy. There are articles on ethics, existentialism, the philosophy of religion, social and political philosophy, and one that brings together the philosophy of emotion and the philosophy of language. The first two articles in this issue are the award winning graduate and undergraduate papers presented at the 52st Annual Meeting of the Florida Philosophical Association, which was held on November 10-12, 2006, at the University of South Florida in Tampa. The presidential address presented at this meeting was "Fifty-Two Card Pick Up" by Greg Ray, Professor of Philosophy at the University of Florida. Professor Ray's address was a humorous performance rather than a traditional paper and thus could not be included here. Thanks go to Shelley Park, who served as site coordinator, and Greg Ray and Sally Ferguson, who served as program coordinators for this event.

This year's award for Outstanding Graduate Student Paper went to Chris Zarpentine of the Florida State University for his paper "Michael Smith, Rationalism, and the Moral Psychology of Psychopathy." Zarpentine argues against the moral rationalists' substantive claim that "there are requirements of rationality … corresponding to the various moral requirements" which are reasons for action. Zarpentine's goal is to undermine this substantive claim using an understanding of psychopathy to justify the position that it is possible we should look for some other foundation in ethics. Psychopaths seem to suffer no rational defect, but instead suffer from an affective deficiency. This results in the psychopath being "fully rational, and yet their moral judgments do not (and will not) converge on the same normative reasons as the rest of us."

Mindi Torrey of the University of South Florida was the recipient of the 2006 Edith and Gerrit Schipper Award for Outstanding Undergraduate Paper for her work "Authenticity: Existential Virtue or Platonic Ideal?" In this work Torrey argues that it is not in more recent continental philosophy that one finds a thorough examination of authenticity, but rather in Plato's dialogues. Plato's account is both comprehensive and critical, according to Torrey, and while showing similarities with the discussion of authenticity found in the "existentialist" writings of Kierkegaard, Heidegger, Sartre, de Beauvoir, and Taylor, the authentic mode of being is most explicitly and successfully expressed by Socrates. Through the maieutic method that leads to self-inquiry one is led to cultivate an authentic life. Torrey also explains that an important feature of the authentic mode of being exhibited by Socrates is the opposition to the contemporaneous power structure. After this consideration of the ideal individual mode of being, the next three articles examine issues in social and political philosophy.

Nathan Andersen, Associate Professor at Eckerd College, takes up the subject of the true community in his article "Hegel on Community and Conflict." Andersen argues that a careful

analysis of the phenomenon of conscientious action in Hegel's *Phenomenology of Spirit* can have profound philosophical and political consequences when considering the possibility of community in our multicultural world. Andersen structures his account on the three stages or "paradigmatic types of relationship that can be set up between conscientious agents," two of which are illustrated through a reading of Jacobi's novel *Edward Allwill's Collection of Letters*. Following Hegel's complex argument, genuine community is neither pre-given nor does it come about through "the mutual admiration of different individuals or groups." Instead, it can arise only through the process of reconciliation or forgiveness, in which the differences and conflicts in our shared situation are recognized as inevitable but productive for the formation of the universal "we" or the "Spirit considered as the community that has come to understand itself."

Joseph Grcic, Professor of Philosophy at Indiana State University, also contributes to the focus on political philosophy in this issue with his article "The Electoral College and Democratic Equality." Here, Grcic argues that the electoral college, not being required to justify its decisions or to make the rationality of its vote public, nor to vote based on the majority vote of citizens, violates the Rawlsian idea of a stable society. Because a stable society and public reason require public discourse and decisions that are consistent with the basic structure of a society based on a Rawlsian theory of justice, the electoral college is unjustified.

James E. Roper, Professor of Philosophy at Michigan State University, continues the discussion of political philosophy with his examination of "Values as a Political Metaframe." Roper's position, critical of the use of simple statements of value in place of complex statements of principle, indicts the use of dichotomous thinking that results in political climates like those created by the Bush Administration. Roper contends that expressions of value alone allow for no balancing of competing principles, and as such, value may override principle. For example, a value like "national security" needs to be understood on the basis of a principle and not only as a value since in the latter case, national security may conflict with and even override other principles of significant import. Roper thus argues that principles are more appropriate and effective in expressing political goals than mere value statements.

In the next article of this issue we turn to the philosophy of religion, as Robert Bass, Assistant Professor of Philosophy at Coastal Carolina University, focuses his inquiry on "Omniscience and the Identification Problem." Of the many puzzles occasioned by a consideration of God's omniscience, Bass focuses on the particular puzzle involving propositional knowledge and argues that a being who knows all the facts about the world is a logical impossibility. In developing this argument, Bass discusses the nature of knowledge and omniscience alone (i.e., of an otherwise unremarkable being referred to as "Todd") before turning to the identification problem, which questions how an omniscient being would know that it is omniscient. According to Bass, who also

carefully considers two possible objections to his argument, no considerations, such as the ontological argument, can resolve the question of omniscience and the identification problem.

The seventh article in this issue by Scott Kimbrough, Associate Professor of Philosophy at Jacksonville University, brings together a consideration of the philosophy of emotion and the philosophy of language. Selected from the 2006 FPA program, "Philosophy of Emotion and Ordinary Language," questions the nature of emotion and the role of judgment in defining emotion. Kimbrough argues against Jesse Prinz's non-cognitive theory of emotion, and in so doing he utilizes P. F. Strawson's ordinary language philosophy, in particular "the distinction between the participant attitude and the objective attitude." On Strawson's account, the participant attitude gives rise to "reactive attitudes" (e.g., resentment or anger) that cannot be adequately accounted for by the objective attitude alone. Kimbrough highlights the conflict between Robert Solomon's cognitive theory of emotion and Prinz's non-cognitive theory by considering *Kansas v. Borman*, a legal case from 1988 involving the murder of woman by her angry and jealous boyfriend Borman. Echoing Solomon, Kimbrough shows that Prinz's attempt to account for emotion independent of judgments "prevents us from saying things about people and their emotions that we need to say."

Jesse Unruh, in "Property Rights of Fetuses: Ontological Problems in the Deprivation Argument against Abortion," argues that the deprivation argument fails because it results in circularity. The basic claims of the deprivation argument are that since fetuses have valuable futures and abortion deprives them of a future, and depriving a fetus of something valuable is immoral, it follows that abortion is immoral. Unruh shows, however, that this argument depends on the supposition that fetuses have property rights (and that those who have property rights are moral agents), but since fetuses are not moral agents, there is no compelling reason to believe that fetuses have such rights. Unruh concludes, therefore, that "since the moral agency of fetuses is essentially what the [deprivation] argument is trying to prove the deprivation argument does not achieve its goal; it is question begging."

Finally, we include at the end of this issue papers from a book symposium held at the 2006 FPA discussing *Donald Davidson: Meaning, Truth, Language, and Reality* (OUP, 2005) by Ernest Lepore and Kirk Ludwig. Simone Evnine, Assistant Professor of Philosophy at Miami University, and Piers Rawling, Professor and Chair of Philosophy at the Florida State University, offer their comments, which are followed by a reply by co-author Kirk Ludwig, Professor of Philosophy at the University of Florida, entitled "The Problem with Radical Interpretation."

We would like to thank our authors for all their efforts, which make this, we think, a strong and significant collection of articles. The ways in which the authors in this issue deal both with problems of current social urgency and general intellectual interest should impress the casual reader as well as the professional academic. We are living on the precipice of impending changes in our

society and the world on the whole. A little more than a year from now, Americans' right to effect a peaceful political revolution through the election process will result in a chance to change the world for the better, to rectify glaring and unconscionable errors of the past and present, and to restore the good reputation and justifiable ideals of American society built on principles of freedom of expression, thought, and action and on respect for individuals, groups, and their rights. Authors in this issue take on hard social and intellectual questions and provide reasoned responses to attempt to begin to answer—and to solve—them.

Please note that correspondence with the authors is possible via the email addresses included in the Notes on Contributors section below. Also deserving of our gratitude are the anonymous referees from around the country, who help insure the high quality of the papers published in *FPR*. While we continue to serve the needs of the Florida Philosophical Association, it is clear from the contents of this issue and recent issues that *FPR* has become much more than a regional publication. As the number of national and international submissions to *FPR* has steadily increased over the past two years, the selection process has naturally become more selective. We continue to extend an open call for high quality papers in all areas of philosophical interest. Last, our thanks go to all the visitors of this site and readers of *FPR*.

Nancy Stanlick and Michael Strawser, Editors
August 2007

Michael Smith, Rationalism, and the Moral Psychology of Psychopathy

Chris Zarpentine, *Florida State University*

Moral rationalism has proved seductive for a number of contemporary ethicists.[1] The claim that moral requirements are rational requirements offers a firm and 'ontologically cheap' foundation for moral objectivity. In *The Moral Problem*, Michael Smith does a fine job of sorting through and distinguishing a number of claims that the rationalist might make. Ultimately he argues for three such claims: the *rationalists' conceptual claim*, the *rationalists' substantive claim*, and the *practicality requirement*. Here, on the basis of psychological research on psychopaths, I will argue against the rationalists' substantive claim. In the process I will discuss some of the implications of these arguments, and then I will consider some responses on Smith's behalf. First, however, I will introduce the three claims and discuss the subtle relations between them.

1. The Logical Terrain

Some philosophers have explicitly distinguished between conceptual claims and substantive or empirical claims that a moral rationalist might make.[2] Smith lays out the rationalists' claims as follows:

> *rationalists' conceptual claim*: "our concept of a moral requirement is the concept of a reason for action; a requirement of rationality or reason."
>
> *rationalists' substantive claim*: "there are requirements of rationality or reason corresponding to the various moral requirements."[3]

Smith rightly distinguishes these two claims and points out:

> Even if we accept the rationalists' conceptual claim, we must still go on to defend the rationalists' substantive claim. And, conversely, even if we deny the rationalists' substantive claim, we must still engage with the rationalists' conceptual claim.[4]

The rationalists' conceptual claim simply falls out of the proper analysis of moral terms. On the other hand, the rationalists' substantive claim is, he says, "about the deliverances of the theory of rational action."[5]

As Smith sets up the distinction, it seems that the truth of the conceptual claim is necessary for the truth of the substantive claim. It is not clear that this is correct. For it might be the case that the rationalists' substantive claim is true, and there are rational requirements corresponding to moral requirements, but at the same time the rationalists' conceptual claim is false, and such correspondence is not required for the truth of moral facts.[6] In any event, I think we would do well to keep these two rationalist claims at least *logically* independent. Later, I will argue that there is a weaker link between these two claims.

Smith argues for both claims. For the conceptual claim, he thinks his arguments are decisive. The substantive claim, however, he maintains, is still an open question, because we do not know if we will (or can) come to a moral consensus. But, Smith is optimistic.

The third claim under consideration, the practicality requirement, is entailed by the rationalists' conceptual claim. Smith defines the former as follows:

> *practicality requirement*: "If an agent judges that it is right for her to ϕ in circumstances C, then either she is motivated to ϕ in C or she is practically irrational."[7]

To further clarify this, we will need to know what 'practical irrationality' means. It is clear that Smith intends 'practical irrationality' to include "weakness of the will and other forms of practical unreason."[8] For, an agent who judges that it is right for her to ϕ, just does judge that she would be motivated to ϕ if she were rational. Thus, if she judges so, and still fails to be motivated, she is irrational 'by her own lights'.[9]

Though I have argued that the rationalists' conceptual claim and their substantive claim are logically independent, I will argue that there is a defeasible link between them. Before I can present this argument, I should say a little more about how Smith cashes out the substantive claim in his anti-Humean theory of normative reasons. Smith argues that these substantive requirements of reason are normative reasons. An agent has a normative reason to ϕ in circumstances C, just in case she would ϕ in C if she were fully rational.[10] Something is desirable if we would desire it if we were fully rational. This theory explains Smith's platitude:

C2: If an agent believes that she has a normative reason to ϕ, then she rationally should desire to ϕ.

Now, we should say something about what "fully rational" means. In explicating, Smith follows Bernard Williams' three conditions of "full rationality," with one qualification on the third condition. On Williams' view, an agent is fully rational if and only if:

(1) The agent has no false beliefs.
(2) The agent has all relevant true beliefs.
(3) The agent deliberates correctly.[11]

Smith argues that the third condition fails to capture the role of deliberation in the destruction of old desires and the creation of new desires. On Smith's view, the fully rational agent must deliberate in an attempt to determine whether her desires are *systematically justifiable*.[12] Smith has in mind a process of "reflective equilibrium" in which we "try to decide whether or not some particular underived desire that we have or might have is a desire to do something that is itself non-derivatively desirable."[13] Smith adds this condition because the truth of normative reasons claims, i.e., of substantive claims, depends on the convergence of the hypothetical desires of fully rational agents in the various circumstances in which they might find themselves.[14]

However, rational amoralists seem to threaten this. They are (apparently) not suffering from any rational defect, and yet they simply do not care about the (alleged) requirements of morality. Thus, we cannot expect their desires to converge with the desires of other fully rational individuals. This would threaten the rationalists' substantive claim that there are normative reasons which are rational requirements.

Now, if it can be shown that the rationalists' substantive claim is false, then we have a defeasible reason for rejecting the rationalists' conceptual claim as well. My argument, then, is this. It is a platitude that many, or perhaps most, adult humans are moral agents, and that there are genuine moral requirements. Thus, our rejection of the substantive claim would seem to lead us toward the view that there are no substantive truths of morality and toward skepticism about human moral agency. Of course, we should allow the *possibility* that J. L. Mackie may be right; however, we should not accept his conclusion lightly.[15] The platitudes that many humans are capable of moral agency and that there are genuine moral requirements, are at least as plausible as the theoretical platitudes which lead to the rationalists' two claims. If we must choose, then we should take seriously the possibility that we need to reject moral rationalism in order to save morality and our

status as moral agents. Thus, if the arguments presented here against the rationalists' substantive claim are sound, then it will put pressure on us to reject the conceptual claim as well. We would do well to consider alternatives to moral rationalism in order to preserve morality.

With these preliminaries out of the way, I will now turn my attention to arguing against the rationalists' substantive claim. As I stated, the existence of rational amoralists would seem to threaten the rationalists' substantive claim. Contemporary examples of such individuals often include psychopaths. Of course, whether we should think of these individuals as actual rational amoralists depends on giving the proper description of such individuals. Thus, it is issues in moral psychology to which we now turn.

2. Moral Psychology and Psychopathy

It is plausible, as Smith states, that it is a platitude that there is an important link between moral judgment or evaluation and motivation. This is captured by Smith's statement:

> What is at issue is thus the very coherence of the idea that deliberation on the basis of our values is practical in its issue to just the extent that it is. As is perhaps evident, this is just the moral problem all over again, redescribed so as to make it explicit that what is at issue is the connection between reason and motivation.[16]

In 1979, Michael Stocker argued that this problem is due to inadequate theories of moral psychology and an under appreciation of common psychological conditions.[17] He argues that:

> motivation and evaluation do not stand in a simple and direct relation to each other, as so often supposed. Rather, they are interrelated in various and complex ways, and their interrelations are mediated by large arrays of complex psychic structures, such as mood, energy, and interest.[18]

Stocker lists a number of "depressions" which can interrupt the normal connection between evaluation and motivation, including "spiritual or physical tiredness," "general apathy," "accidie," "thorough despair," and so on. Another strain of individuals which demonstrate an abnormal connection between evaluation and motivation involve conditions like psychopathy.[19] Recent psychological research has helped to shed some light on these psychological phenomena. What is particularly interesting about all of these conditions is that such individuals are notably *disaffected*, i.e.,

they lack normal emotional responses in certain circumstances. In this section, I will argue that a proper understanding of psychopathy, and specifically the role of affect in mediating the relation between evaluation and motivation, undermines the rationalists' substantive claim.[20]

Psychopaths are problematic because they apparently suffer from no rational defect, and yet they seem to be morally lacking in an important sense. One immediate response is that psychopaths do not make real moral judgments, but only make judgments in the "inverted commas" sense.[21] To see that this response misses the point, we must appreciate that this is a conceptual claim. Thus, it must be evaluated by the standards of Smith's conceptual claim.[22] Recall that Smith's conceptual claim falls out of the proper analysis of moral terms. This analysis is based on the set of platitudes that we accept when we come to have mastery of a concept.[23] However, it does seem to be a platitude that psychopaths *do* make moral judgments, in the full sense. Recent research seems to support this.[24]

Despite the fact that we think psychopaths make moral judgments, it should be clear that their judgments do differ in some important ways from the judgments of normals. If such divergence is not due to any rational defect, then the moral rationalist is in trouble. The fact that the moral behavior and motivation of psychopaths often diverges so radically from that of normals suggests that their moral judgments are significantly different from ours. Understanding what is wrong with psychopaths will help us to understand the relation between moral judgment and motivation. As I suggested above, this relation is mediated by an affective mechanism.

There has been recent research by R. James Blair and others on psychopathic criminals in English prisons.[25] As a control group, Blair studied non-psychopathic criminals. Blair utilized R.D. Hare's Revised Psychopathy Checklist to distinguish psychopathic from non-psychopathic criminals.[26] His research presents interesting findings. Blair finds that psychopaths fail to distinguish between *moral transgressions* and *conventional transgressions*. Other research, demonstrates that psychopaths have deficient *affective responses* to thoughts or images of distress in other individuals.[27] I suggest that this deficiency in affective responses explains the failure of psychopaths to appreciate the moral/conventional distinction, and furthermore that this affective mechanism explains the link between moral judgment and motivation in normal individuals. First, we should explain these two findings in a little more detail.

Blair explains the moral/conventional distinction as follows:

> Moral transgressions have been defined by their consequences for the rights and welfare of others, and social conventional transgressions have been defined as violations of behavioral uniformities that structure social interactions within social systems.[28]

Subjects in Blair's study were presented with scenarios of transgressions, and then asked to judge on three conditions: permissibility, seriousness, and authority jurisdiction. Blair's study concluded that "while non-psychopaths made the moral/conventional distinction, the psychopaths did not."[29] This finding suggests that despite the fact that psychopaths make moral judgments, they are unable to appreciate the distinct nature of moral judgments as opposed to conventional judgments.

This could be explained by the fact that psychopaths lack normal affective responses to distressing situations, such as those involved in moral transgressions. Blair *et al.* reports that "studies have shown that psychopaths do not show arousal responses to the distress of others."[30] This is measured in part by their physiological responses to distressing images or thoughts, but can also be measured indirectly by examining their behavioral tendencies. Thus, it would seem that that an affective deficiency—not a rational one—is responsible for the distinct nature of psychopathic moral judgments. This in turn gives rise to their peculiarly amoral behavior.

It is interesting that Philippa Foot, in arguing against the view that moral considerations have "automatic reason-giving force," asks:

> Why cannot the indifferent amoral man say that for him [the moral "should"] gives no reason for acting, treating [the moral "should"] as most of us treat [the "should" of etiquette]?[31]

This seems to be precisely how psychopaths do think of the moral "should," simply as a kind of conventional "should." In attempting to explain how exactly the moral "should" differs from the "should" of etiquette she hypothesizes that it might have to do with the fact that moral rules are enforced more strictly than rules of etiquette.[32] We do, however, have good evidence that affect is very important in mediating the relation between moral judgment and motivation.

What is important at present is that the deficiency of psychopaths is not due to any rational defect. It is due to an affective deficiency. Thus, some psychopaths may be fully rational, and yet their moral judgments do not (and will not) converge on the same normative reasons as the rest of us. The truth, however, of the rationalists' substantive claim depends on the agreement of all fully rational individuals on the same set of normative reasons. The existence of rational psychopaths seems to demonstrate the falsity of the rationalists' substantive claim. As long as there are rational amoralists like the psychopath, there will not be convergence.

3. Replies on Behalf of Smith

There are a number of responses Smith might make, two of which I will consider here. I will argue that neither is successful. First, Smith might argue that psychopaths fail to have *systematically justifiable desires*. Remember that Smith added this qualification to Williams' analysis of "fully rational." Here we have been assuming that at least some psychopaths are capable of being fully rational, in the sense of having no false beliefs, all relevant true beliefs, and deliberating correctly. Remember, also, that to have a set of systematically justifiable desires, we must undergo a process of reflective equilibrium in which we see if "our specific value judgments would be more satisfyingly justified and explained by seeing them as all falling under a more general principle."[33] This may, in turn, give rise to new desires. "For," as Smith says, "an evaluative belief is simply a belief about what would be desired if we were fully rational, and the new desire is acquired precisely because it is believed to be required for us to be rational."[34]

However, we need to keep in mind that having systematically justifiable desires is part of Smith's analysis of what it is to be "fully rational." Thus, full rationality cannot require having all the relevant true *evaluative* beliefs, for that would make his account viciously circular. Such an account would presuppose exactly what is at issue: whether there are, in fact, any true beliefs of the form "ϕ would be desired if S were fully rational." Such beliefs are supposed to be generated by Smith's account of "full rationality"; they cannot be presupposed by it. Thus, Smith cannot argue that the desires of psychopaths fail to be systematically justified, for they very well may be, even if they diverge from the desires of normals.

Now if this is correct, then on what basis may the psychopaths' desires be rationally criticized? This leads us to Smith's second response. In response to similar arguments by Philippa Foot and Gilbert Harman involving rational amoralists, Smith levels the charge of intellectual arrogance.[35] Smith might allege that the psychopath "can give no account of why his own opinion about what fully rational creatures would want should be privileged over the opinion of others; he can give no account of why his opinion should be right, others' opinions should be wrong."[36] Smith's point is not that the psychopath is wrong because he disagrees with *us*, but rather that "he rejects the very idea that the folk possess between them a stock of wisdom about such matters… And yet, ultimately, this is the only court of appeal there is for claims about what we have normative reason to do."[37] The psychopath "sticks with this opinion despite the fact that virtually everyone disagrees with him" and "he does so without good reason."[38]

This response, however, seems to suppose that convergence is forthcoming. But as Smith himself states:

> The mere fact that a convergence in the hypothetical desires of fully rational creatures is required for the truth of normative reason claims does nothing to guarantee that such a convergence is forthcoming….[39]

Smith grants that it is a logical possibility that "no systematic justification of our desires is forthcoming," but that "it seems more reasonable to think that such a justification is forthcoming."[40] But neither does it seem *irrational* or *unreasonable* to think that a systematic justification which would lead to convergence is not forthcoming. If this is the case, then the psychopath is not "without good reason."

Furthermore, even if there were such a convergence, it would not be clear that this would be sufficient to establish the objectivity of morality in rational requirements. Such a convergence might be due to the influence of non-rational factors.[41] As Richard Rorty put it:

> If it seems that most of the work of changing moral intuitions is being done by manipulating our feelings rather than by increasing our knowledge, that is a reason to think there is no knowledge of the sort that philosophers like Plato, Aquinas, and Kant hoped to get.[42]

Smith makes much of the resolution of disagreement over "slavery, worker's rights, women's rights, democracy and the like."[43] He sees this as the triumph of reason: "We must not forget that there has been considerable moral progress."[44] But to call these resolutions "moral progress" is to assume that convergence is always progressive, something we may not be entitled to assume. Rorty concludes that the past two hundred years of moral progress:

> are more easily understood not as a period of deepening understanding of the nature of rationality or of morality, but rather as one in which there occurred an astonishingly rapid progress of sentiments, in which it has become much easier for us to be moved to action by sad and sentimental stories.[45]

Thus, the rational amoralist may have good reason to think that either convergence is not forthcoming, or that it is not indicative of progress and objectivity.[46] Certainly such considerations give reasonable weight to the rational amoralists' skepticism about normative reasons. As long as such skepticism is indeed reasonable, then Smith cannot level the charge of intellectual arrogance.

4. Conclusion

I have argued that the rationalists' substantive claim is flawed. If these arguments are sound, this puts pressure on us to reject the rationalists' conceptual claim, or else accept the view that there are no substantive truths of morality. In addition, since the practicality requirement is itself entailed by the rationalists' conceptual claim, if we wish to maintain it, we will need to find an independent argument for it. However, it seems that similar evidence relating to psychopaths would undermine the practicality requirement as well. Perhaps we do not need to reject moral rationalism outright; perhaps a weaker version of the thesis may be defended. Nevertheless, as I have suggested earlier, we may do well to look elsewhere for a foundation in ethics. And any plausible moral theory will have to take into consideration the kind of issues in moral psychology which I have discussed here.[47]

Notes

[1] For example: Thomas Nagel, *The Possibility of Altruism* (Princeton: Princeton UP, 1970); Peter Singer, *How Are We to Live* (Amherst: Prometheus Books, 1995); and Michael Smith, *The Moral Problem* (Oxford: Blackwell Publishing, 1994).

[2] For example: Shaun Nichols, *Sentimental Rules* (Oxford: Oxford UP, 2004) and Smith, *op. cit.*

[3] Smith 64-5.

[4] Smith 64.

[5] Smith 65. It is precisely this distinction which allows J. L. Mackie to argue for his 'error theory'; see his *Ethics: Inventing Right and Wrong* (London: Penguin, 1977). Roughly speaking, his argument consists of the conjunction of a hypothetical form of the rationalists' conceptual claim and the negation of the rationalists' substantive claim. His conclusion is thus that there are no such moral requirements.

[6] For example, a Platonic conception of morality seems to be consistent with the rationalists' substantive claim. On this view, moral facts are eternal truths in some supernatural realm. Nevertheless, a plausible moral epistemology might maintain that reason is the process by which we become aware of such truths.

[7] Smith, 61.

[8] Smith 61

[9] Smith 62. It should be noted that one way of arguing against the rationalist is to deny the practicality requirement. Doing so would undermine the rationalists' conceptual claim, and thus—at least on Smith's view—would undermine the substantial claim, though I have argued that it is not clear that the substantive claim requires the truth of the conceptual claim, even for the former to be an interesting claim in its own right. David O. Brink and other 'externalists' take up this strategy, and Smith attempts to defuse these criticisms. I think that a better strategy—and the one I will pursue here—is to attack the rationalists' substantive claim directly. See David O. Brink, *Moral Realism and the Foundation of Ethics* (Cambridge: Cambridge UP, 1989).

[10] Smith 151.

[11] Smith 156. For Williams' view see his "Internal and External Reasons," reprinted in *Moral Luck* (Cambridge: Cambridge UP, 1980).

[12] Smith 158-9.

[13] Smith 159.

[14] Smith 173.

[15] Mackie, *op. cit.*

[16] Smith 137.

[17] Michael Stocker, "Desiring the Bad: An Essay in Moral Psychology," *Journal of Philosophy* 76 (1979): 738-753.

[18] Stocker 739.

[19] There is a similar condition, so-called "acquired sociopathy" which involves damage to the prefrontal cortex, and produces similar behaviors as psychopathy. See A. Bechara, H. Damasio, and A. R. Damasio, "Emotion, Decision Making and the Orbirofrontal Cortex," *Cerebral Cortex* 10 (2000): 295-307; R. James Blair and L. Cipolotti, "Impaired Social Response Reversal: A Case of 'Acquired Sociopathy'." *Brain* 123 (2000): 1122-1141; A. R. Damasio, D. Tranel, and H. Damasio, "Individuals with Sociopathic Behavior Caused by Frontal Damage Fail to Respond Autonomically to Social Stimuli." *Behavior and Brain Research* 41 (1990): 81-94; and A. R. Damasio *Descartes' Error*. (New York: Grosset/Putnam, 1994) for more on this condition, which itself presents interesting and relevant empirical results to contemporary moral psychology. See Adina Roskies, "Are Ethical Judgments Intrinsically Motivational? Lessons from 'Acquired Sociopathy'," *Philosophical Psychology* 16 (2003): 51-66, for one application of these results to 'belief-internalism'.

[20] We could use a hypothetical rational amoralist, like Hume's sensible knave or Gilbert Harman's "successful criminal" to argue against the rationalists' substantive claim, see David Hume, *A Treatise of Human Nature* (Oxford: Clarendon P, 1968) and Harman, "Is there a Single True Morality?" in David Copp and David Zimmerman, eds., *Morality, Reason and Truth* (Totowa: Rowman and Allanheld, 1985). However, the use of actual individuals makes the problem more pressing for the rationalist. Furthermore, it highlights the crucial role that affect plays in moral psychology, a fact that any successful moral theory will have to account for.

[21] See R.M. Hare, *The Language of Morals* (Oxford: Oxford UP, 1952).

[22] This point is made nicely by Nichols *op. cit.*

[23] Smith 31.

[24] Nichols *op. cit.* notes: "If it is a platitude that psychopaths really make moral judgments, it will be difficult to prove that conceptual rationalism captures the folk platitudes surrounding moral judgment" (75). He continues that the "inverted-commas enthusiast" might "maintain that a process of reflective equilibrium would lead people to reject the platitude about psychopathic moral judgment" (75). But this claim is an empirical one, and at present an unsubstantiated one. Since I am primarily interested in addressing the rationalists' substantive claim, I will have to leave this issue as it stands.

[25] See R. James Blair, "A Cognitive Developmental Approach to Morality: Investigating the Psychopath" *Cognition* 57 (1995): 1-29; and R. James Blair, J. Jones, F. Clark, and M. Smith, "Is the Psychopath 'Morally Insane?'" *Personality and Individual Differences* 27 (1995): 135-145.

[26] R.D. Hare, *The Hare Psychopathy Checklist—Revised* (Toronto: Multi-Health Systems, 1991).

[27] A.S. Aniskiewicz, "Autonomic Components of Vicarious Conditioning and Psychopathy" *Journal of Clinical Psychology* 35 (1979): 60-67; and T.H. House and W.L. Milligan, "Autonomic responses to modeled distress in prison psychopaths," *Journal of Personality and Social Psychology*, 34 (1976): 556-560.

[28] Blair 5.

[29] Blair 20.

[30] Blair *et al.*, 750. They cite Aniskiewicz and House and Milligan *op. cit.*

[31] Philippa Foot, "Morality as a System of Hypothetical Imperatives," *Philosophical Review* 81 (1972): 305-316. This quotation does not appear in the original published version. It is an addition to a footnote published in a slightly revised version of the paper in Steven M. Cahn and Joram G. Haber, eds., *20th Century Ethical Theory* (Englewood Cliffs: Prentice Hall, 1995).

[32] Foot asks, rhetorically:

> But are we then to say that there is nothing behind the idea that moral judgments are categorical imperatives but the relative stringency of our moral teaching? I believe that this may have more to with the matter than the defenders of the categorical imperative would like to admit. For if we look at the kind of thing that is said in its defense we may find ourselves puzzled about what the words can even mean unless we connect them with the *feelings* that this stringent teaching implants (310, *emphasis added*).

Indeed, Foot seems to be on the right track. It does seem to be the *feelings* that we associate with moral judgments which makes them distinct. And it is precisely these feelings which psychopaths are missing as a result of their affective deficiency. Certainly teaching has a lot to do with the proper development of this faculty; however, psychopaths seem to be deficient in a way that precludes the proper development of this capacity. Of course, moral development plays an important part in the mediation of moral judgment and motivation, but such an investigation at present would take us too far a field.

[33] Smith 160.

[34] Smith 160.

[35] See Foot and Harman *op. cit.*

[36] Smith 195.

[37] Smith 195-6.

[38] Smith 195.

[39] Smith 173.

[40] Smith 201.

[41] A related worry might be that any putative convergence of normative reasons is not due to rationality, but systematically biased human intuitions. We may indeed converge in our desires, but such convergence does not guarantee objectivity. Such convergence may be a product of our systematically biased faculties, much the same way in which our *prima facie* judgments of optical illusions converge but are nonetheless flawed. Smith could respond that reason allows us to correct such systematic biases, just as further investigation leads us to correct our judgments about optical illusions. However, even this cannot guarantee that our judgments about our moral obligations to the distant needy or to non-human animals will converge on the *true* normative reasons.

[42] Richard Rorty, "Human Rights, Rationality, and Sentimentality," in his *Truth and Progress: Philosophical Papers*, vol. 3 (Cambridge: Cambridge UP, 1998), 172.

[43] Smith 188.

[44] Smith 188.

[45] Rorty 185.

[46] Indeed, if I am right about the role of affect in moral judgment, this provides yet another reason to think that convergence needn't indicate objectivity.

[47] Special thanks to David McNaughton and Clifford Sosis for helpful discussion.

Works Cited

Aniskiewicz, A.S. "Autonomic Components of Vicarious Conditioning and Psychopathy." *Journal of Clinical Psychology* 35 (1979): 60-67.

Bechara, A., Damasio, H., & Damasio, A. R. "Emotion, Decision Making and the Orbirofrontal Cortex." *Cerebral Cortex* 10 (2000): 295-307.

Blair, R. James. "A Cognitive Developmental Approach to Morality: Investigating the Psychopath." *Cognition* 57 (1995): 1-29.

Blair, R. James, L. Jones F. Clark, and M. Smith. "Is the Psychopath 'Morally Insane?'" *Personality and Individual Differences* 27 (1995): 135-145.

Blair, R. James and L. Cipolotti. "Impaired Social Response Reversal: A Case of 'Acquired Sociopathy'." *Brain* 123 (2000): 1122-1141.

Brink, David O. *Moral Realism and the Foundation of Ethics*. Cambridge: Cambridge UP, 1989.

Cahn, Steven M. and Joram G. Haber. *20th Century Ethical Theory*. Englewood Cliffs, NJ: Prentice Hall, 1995.

Damasio, A. R., Tranel, D. & Damasio, H. "Individuals with Sociopathic Behavior Caused by Frontal Damage Fail to Respond Autonomically to Social Stimuli." *Behavior and Brain Research* 41 (1990): 81-94.

Damasio, A. R. *Descartes' Error*. New York: Grosset/Putnam, 1994.

Foot, Philippa. "Morality as a System of Hypothetical Imperatives." *Philosophical Review* 81 (1972): 305-316; revised version reprinted in Cahn and Haber, eds. *20th Century Ethical Theory*.

Hare, R.D. *The Hare Psychopathy Checklist—Revised*. Toronto: Multi-Health Systems, 1991.

Hare, R.M. *The Language of Morals*. Oxford: Oxford UP, 1952.

Harman, Gilbert. "Is There a Single True Morality?" In *Morality, Reason and Truth*. David Copp and David Zimmerman, eds. Totowa: Rowman and Allanheld, 1985.

House, T.H. and W.L. Milligan. "Autonomic responses to modeled distress in prison psychopaths." *Journal of Personality and Social Psychology*, 34 (1976): 556-560.

Hume, David. *A Treatise of Human Nature*. Oxford: Clarendon P, 1968.

Mackie, John L. *Ethics*: *Inventing Right and Wrong*. London: Penguin, 1977.

Nagel, Thomas. *The Possibility of Altruism*. Princeton: Princeton UP, 1970.

Nichols, Shaun. *Sentimental Rules*. Oxford: Oxford UP, 2004.

Rorty, Richard. "Human Rights, Rationality, and Sentimentality." In *Truth and Progress: Philosophical papers*, vol. 3. Cambridge: Cambridge UP, 1998.

Roskies, Adina. "Are Ethical Judgments Intrinsically Motivational? Lessons from 'Acquired Sociopathy'." *Philosophical Psychology* 16 (2003): 51-66.

Singer, Peter. *How Are We to Live*. Amherst: Prometheus Books, 1995.

Smith, Michael. *The Moral Problem*. Oxford: Blackwell Publishing, 1994.

Stocker, Michael. "Desiring the Bad: An Essay in Moral Psychology." *Journal of Philosophy*, 76 (1979): 738-753.

Williams, Bernard. "Internal and External Reasons." Reprinted in *Moral Luck*. Cambridge: Cambridge UP, 1981.

Authenticity: Existential Virtue or Platonic Ideal?

Winner of the Gerritt and Edith Schipper Undergraduate Award
for Outstanding Undergraduate Paper at the
52nd Annual Meeting of the Florida Philosophical Association

Mindi Torrey, *University of South Florida*

Authenticity is commonly identified as a uniquely existential matter and generally as an existential virtue. Both designations, that of being uniquely existential and that of being an existential virtue, require inquiry. The current discussions on authenticity extend from continental philosophy; however, I argue in this paper that along with countless metaphysical and epistemological topics, the notion of authenticity has been a philosophic matter since Plato wrote his dialogues, and has yet to be presented as comprehensively or as successfully. I present here a conception of authenticity from a perspective supported by a contemporary reading of Plato following a brief description of some of the current existential representations of authenticity. In addition, I offer parallels between popular notions of authenticity and the elements deemed critical in its definition, with accounts revealed in Plato's dialogues representing the ideal of authenticity. My objective is to show that the ancient account includes critical elements deficient in the current definitions of authenticity.

Marjorie Greene describes authenticity as an exclusively existential virtue: "What the existentialist admires is not the happiness of a man's life, the goodness of his disposition, or the rightness of his acts but the authenticity of his existence. This is, I think, the unique contribution of existentialism to ethical theory."[1] Plato's discussions of various themes in his dialogues, including those that Greene asserts are of no concern to the existentialist, are embedded in a complex structure that is itself a dynamic and recurring demonstration of authenticity. Plato's dialogues as a body of work represent an exhaustive argument for the necessity of authenticity; in fact it is the most paramount pursuit fundamental to rational beings. Many of the existentialist arguments bear a striking resemblance to the views on authenticity introduced in the Platonic dialogues yet none is as comprehensive, successful or enduring.

Sartre asserts his version of authenticity as a moral imperative based in natural freedom. An agent acts with regard to the realization of her freedom; she is aware of her limitations, the ambiguity of her freedom. However, "If value is totally contingent upon freedom, then such talk of

ethical 'right' and 'wrong' becomes inappropriate since it implies an ethical objectivity totally lacking in Sartre's theory."[2] Sartre's authenticity seems more of a contradiction of reasoning than an existential virtue. De Beauvoir discusses authenticity as an ambiguous existence requiring the willing of herself "to be a disclosure of being" in order to become authentic, asserting her freedom.[3] That an agent wills herself as a disclosure of being seems to apply more to epistemology than to ethics. There is no presence of a value in de Beauvoir's account of authenticity. Therefore, like Sartre's, ethical objectivity is lacking.

For Kierkegaard, meaning in life is reliant on a commitment to unity. The conflict between the temporal and the eternal aspects of our lives cannot be resolved with rational dialectic. The agent, according to Kierkegaard, achieves unity only by making a continuously renewed choice, a commitment to another person or being. Only such a person aimed at achieving a coherent unity can be an existing individual. Kierkegaard suggests in *Concluding Unscientific Postscript*, "existence itself, the act of existing is a striving."[4] Thus, the meaning in life, according to Kierkegaard, relies on the synthesis of the temporal and the eternal through a continuously restored commitment to an external being.

Martin Heidegger and Charles Taylor have theories of authenticity that are distinctive, if similar. Both require self inquiry, Heidegger through *Dasein* and Taylor through self interpretation. Both view authenticity as a mode of being or a continual process of being in the world. Taylor requires for authenticity a need for "openness to horizons of significance." Heidegger relies on an anti-subjectivist standpoint: we do not create the significance that is our facticity, yet it is something in which we are inherently involved.[5] Both believe that authenticity is a continuously recurring process although we cannot find a "final escape from our original social position."[6] The main distinction between Taylor and Heidegger is that Taylor's version of authenticity is normative. Heidegger maintains that an inauthentic mode of being does not reduce the value since values are "those things belonging to the very 'thy-self' that authenticity is meant to overcome."[7] Yet, as Gerry Stahl notes, "Heidegger shirks responsibility for the claim inherent in the word 'authenticity' to be presenting a positive doctrine of the good life when he insists that he is using the word as a value-free technical term."[8] It seems that the uniquely existential "virtue of authenticity" has confounded the existentialists themselves as to whether authenticity is in fact a virtue at all.

Alexander Nehamas' book *Virtues of Authenticity: Essays on Plato and Socrates* presents various essays dealing with the Platonic dialogues. I was initially perplexed that in the many pages of a book of essays about Plato and Socrates entitled "Virtues of Authenticity" the word "authenticity" shows up only in the introduction; it is not found in any of the essays themselves.[9] Similarly, while Plato does not set forth an argument per se for authenticity, it is the theme and the objective that he and

Socrates bear within the complex system of the dialogues, expecting the interlocutors and the readers alike to experience authenticity, as Heidegger and Taylor suggest, through self-inquiry.

There is much debate regarding virtue in Plato's writing; there is scarcely a topic more discussed than the nature of virtue in the dialogues. Nehamas suggests that in order to be a teacher of *arete* one would have to either know or claim to know what comprises the good life, or one would have to set oneself up as an example of what it is to lead the good life. Socrates, Nehamas suggests, fulfills neither of these requirements.[10] To limit the question of the good life to whether Socrates claimed to know or to be an example of virtue misses the point, for it seems obvious that Plato did not write the dialogues with a primary goal of communicating or making clear his proposal of an objective moral treatise. Plato's scheme is an involved multilayered set of works presenting Socrates not as a teacher, but rather as a midwife. While Plato's admiration of Socrates as a virtuous man is evident in his writings, virtuosity is not the isolated function of Socrates as a character in Plato's works. Socrates functions not as a teacher of virtue, but rather as a midwife of authenticity.

Heidegger's authenticity involves *Dasein*, a temporal account of a mode of being not as a decidedly achieved state, but rather as a continual process through which one embraces life. Similarly, Charles Taylor's authenticity requires that it "is not a thing or an objectified state; rather, it is a continual process of dialogue with the social world."[11] Plato illustrates this very aspect of authenticity not only in Socrates' singularly famous expression "the unexamined life is not worth living," but throughout his dialogues with the presentation of Socratic Inquiry.

In the *Protagoras*, we see Socrates engaged in dialogue with Protagoras, a popular sophist. Responding to the sophist's insistence that being in command of poetry is the "greatest part of a man's education," Socrates illustrates, through Socratic Inquiry, both that Protagoras is claiming knowledge about a subject he has not fully considered, and that being and becoming are two different ways of conceiving existence. Plato uses Socrates' character here to reveal to his readers and to Socrates' audience the value of continual reflection in contrast to the objectified state of knowledge that Protagoras celebrates. Socrates makes this point with a playful gloss on Protagoras' explication of Simonides' poem. It is not hard for a man to *become* good, not truly good, but truly hard. But to *be* good, Socrates says, it is impossible. A man who becomes good is not always good, but sometimes good and sometimes bad since a man can be "thrown down"; yet it is the duty for that man to continuously strive to become good.[12] In the presence of a large crowd of impressionable Athenian young men, Socrates contests Protagoras' claims of knowledge by illustrating to the audiences in and of the dialogue that what exists in the social world must necessarily be questioned. Throughout the dialogues Socrates impresses upon both audiences the

importance of Socratic Inquiry. Undermining traditional discourse is imperative to the creation of authenticity.

In response to the charge that he has corrupted the youth, in the *Apology*, Socrates states that he has "never promised to teach them anything and [I] have not done so. If anyone says that he has learned anything from me…be assured that he is not telling the truth."[13] This is important not only in distinguishing Socrates from the sophists, but in establishing Plato's ideal of authenticity. Socrates, like Plato desires not to teach others what he claims to know, but rather to influence others to pursue self-knowledge through inquiry. Authenticity is a process of both questioning social discourse and gaining self-awareness. Socrates says of the sophists, "try asking one of them something, and they will be as unable to answer your question or to ask one of their own as a book would be. Question the least little thing in their speeches and they will go on like bronze bowls that keep ringing for a long time after they have been struck and prolong the sound indefinitely unless you dampen them."[14]

Kierkegaard defines truth as "*An objective uncertainty held fast in an appropriation-process of the most passionate inwardness*;" this, he says, is "the highest truth attainable for an *existing* individual."[15] Kierkegaard himself points out that Socrates contributed to the "passion of the infinite" in his struggle with uncertainty. Kierkegaard writes:

> But Socrates! He puts the question objectively in a problematic manner: *if* there is an immortality. …On this "if" he risks his entire life, he has the courage to meet death, and he has with the passion of the infinite so determined the pattern of his life that it must be found acceptable—*if* there is an immortality.[16]

Plato's Socrates displays a "passion of inwardness" inherent in the Socratic ignorance which exemplifies Kierkegaard's belief that the truth is related to an existing individual. Kierkegaard suggests that the subjective truth offered in the dialogues through the unfolding of Socratic ignorance may offer more truth than the "entire objective truth of the System, which flirts with what the times demand."[17]

In the *Meno,* Socrates suggests learning is recollection which, by his own admission, is nothing more than a device to get Meno's character to agree that whether we believe ourselves to know or not what it is we are looking for, we should always search:

> I do not insist that my argument is right in all other respects, but I would contend at all costs both in word and deed as far as I could that we will be better men, braver and less idle, if we believe that one must search for the things one does not know, rather than if we believe that it is not possible to find out what we do not know and that we must not look for it.[18]

Plato's Socrates does not rely on his arguments about virtue and justice to inculcate theories or doctrines; rather he relies on his method as a means of assisting others in acquiring what we would term an *authentic mode of being*.

Guignon suggests that on a Platonist reading of Socrates, "humans are regarded as parts of a wider cosmic totality, placeholders in a cosmic web of relations in which what anything is—its *being* as an entity of a particular sort—is determined by its place and function within that wider whole," and that "to know yourself, then is to know above all what your *place* is in the scheme of things—what you are and what you should be as that has been laid out in advance by the cosmic order."[19] Socrates and Plato consistently urge their interlocutors, observers and readers away from this notion of social place holding. It is this very notion that the dialogues attempt to refute. In the Republic I, for instance, prior to the creation of the ideal city-state and the search for justice, Socrates' character announces "don't think that in searching for justice, a thing more valuable than even a large quantity of gold, we'd mindlessly give way to one another or be less than completely serious about finding it. You surely mustn't think that, but rather—as I do—that we're incapable of finding it."[20] Socrates and Plato are not trying to advance a political or social conception of justice through rule by philosopher kings and throwing babies over cliffs. The objective is quite the opposite; to engage the young men of Athens and the reader to inquire and reflect upon individual and personal conceptions of justice, virtue, education, aesthetics, goodness and all that is authentic in the search for subjective truth, despite the backdrop of the "cosmic order." This is the very notion that the Platonist reading of the dialogues disallows. Plato's Socrates realizes that knowledge of justice, virtue or the ideal polis are not attainable, at least in the world of becoming. The Platonic writings are inherently more complicated than a social or political ideology. They present the occasion for authentic awakening through active dialectic. The role of midwife is thus taken up by both Plato and Socrates in encouraging Socratic Inquiry, which induces self-inquiry and leads to authenticity.

A major aspect of authenticity that seems to be absent in many current definitions is that of contesting power through opposition. While most discussions of authenticity include some mention of rejecting social influences in the process of finding a true or genuine mode of becoming, the presence of the power structure is a more complex and pervasive obstacle to authenticity. Plato addressed this force comprehensively. He is not advocating but rather questioning and addressing the "omnipresence of power relations and their tendency to combine and become systematized and institutionalized" which, as Foucault believes, "must be tempered by the fact that power is also 'rooted deep in the social nexus.'" Foucault notes that if there are relations of power throughout every social field it is because there is freedom everywhere and that "people…are much freer than they feel."[21] It is no coincidence that Socrates' interlocutors suggest definitions of justice in terms of the "advantage of the stronger"[22] or of virtue as "to desire beautiful things and have the power to acquire them."[23] Socrates suggests that people are ruled by the power of appearance which "often makes us wander all over the place in confusion, often changing our minds about the same things and regretting our actions and choices with respect to things large and small, the art of measurement in contrast, would make the appearances lose their power by showing us the truth, would give us peace of mind firmly rooted in the truth and would save our life."[24] Plato's comments on the *power of appearance* and the *art of measurement* exemplify the struggle of authenticity against the prevalence of power. The art of measurement, the balancing of the large and immediate desires with the small and distant reasons, leads the individual toward an authentic mode of being. The power of appearance, on the other hand, leads the individual to a position within a significant set of power relations in which her authenticity is constrained.

The ideal of authenticity is an existential concern insofar as it relates to what Teodoros Kiros calls the "existentially serious person." He notes that such a person finds herself in a situation in which "a concern with existence, which is the ultimate freedom of the individual, is subordinated to the increased legally sanctified power of the community to decide on the behalf of the helpless, neurotic, pleasure-seeking, and dangerous individual who is incapable of cautiously determining the desires of the self."[25] Socrates maintains that those who are controlled by the power of appearance are depriving themselves of the ultimate pleasure, of their very "salvation in life" which those who are willing to practice the art of measurement would struggle for.

Plato aims to break down this larger than life non-reflective presence of the power of the community through Socrates' continued inquiries concerning sophistry, education, justice, virtue and knowledge. Socrates makes this claim in the *Apology:*

> Good Sir, you are an Athenian, a citizen of the greatest city with the greatest reputation for both wisdom and power; are you not ashamed of your eagerness to possess as much wealth, reputation and honors as possible, while you do not care for nor give thought to wisdom or truth, or the best possible state of your soul?

Socrates holds resolutely to his purpose of authentic midwifery, continuing, as he faces death,

> if one of you disputes this and says he does care, I shall not let him go at once or leave him, but I shall question him, examine him and test him, and if I do not think he has attained the goodness that he says he has, I shall reproach him because he attaches little importance to the most important things and greater importance to inferior things. [26]

As agents we have the freedom to choose our actions. Plato makes a case for authenticity in the *Republic* suggesting the use of the *art of measurement* as opposed to the *power of appearance*. Socrates says to Glaucon, "he'll go on imitating, even though he doesn't know the good or bad qualities of anything, but what he'll imitate, it seems, is what appears fine or beautiful to the majority of people who know nothing," and "an imitator has no worthwhile knowledge of the things he imitates, that imitation is a kind of game and not something to be taken seriously."

Freedom is available by existence alone and adults cannot be spared this "anguish," but dissention or revolt is available with the subsistence of power. Resistance, however, is "not an external struggle *against* power, but an internal and dyadic *exercise* of power relations, over others as much as over ourselves" and these relations are "sometimes extended by the position of those who are dominated."[27] Socrates puts this freedom to the ultimate test in the *Apology*. With the opportunity to appeal to the emotions of the jurors and defend himself in a way that would allow him to be free from the punishment of the government, Socrates does not waver in his life of continuous struggle toward the ideal of authenticity. Socrates declares that his unpopularity will be his undoing; that it would not be his accusers, but the envy and slander of many and that other good men had been ruined this way and he would not be the last.[28] Socrates is unconcerned that the men of Athens will not understand his method of defense by truth, since justice is not measured by truth, but rather by tears and pity. He is interested in self-truth which revolts from power.

In the *Apology,* Plato takes a firm stance for authenticity. He suggests that the justice wrought through power relations is not justice at all, and that soliciting for rights in such a system is inauthentic. Facing death, Socrates declares "I will not yield to any man contrary to what is right, for fear of death, even if I should die at once for not yielding."[29] Knowing that his defense will be unsuccessful, Socrates chooses to continue to act authentically and is not swayed by the power of the city-state's sentence of death. While we can never exist independent of society, we always have the choice to act responsibly, choosing authenticity and spurning the power which believes itself to have infinite control over every human being. The dialogues elevate Socrates as the ideal authentic self in the sense that Guignon describes the authentic individual to be "an individual who can stand alone, shedding all status relations and social entanglements, in order to immerse him- or herself in 'sheer life'" and the "ultimate form of artistic creation."[30]

Following a discussion of various thoughts on authenticity, Guignon explores the possibility that "it has come to appear that mucking around inside the mental container, far from leading us to a better richer life, might be a path to confusion and despair" and "there is the risk of slipping into a life so prone to self-absorption and self-surveillance that one becomes isolated from all but those who share this preoccupation." He then counters this position stating that despite the "arduous process" involved in authenticity, most people view it as an "ideal character trait or personal virtue that is necessary to living the best possible life for humans under modern circumstances."[31] Taylor's authentic mode of being comprises the struggle against meaninglessness and Heidegger's is a triumph over the triviality of existence. Both believe that authenticity allows for a "world that is far richer, fuller, and more vivid."[32]

It is necessary to note that "it is not enough to merely assert the worth of certain ideals; there must also be a means of sustaining them," and that "high standards need strong sources."[33] Plato's dynamic and enduring dialogues promote, by method of Socratic Inquiry, the repudiation of imitation and unchallenged beliefs. The dialogues are not merely attempts to define authenticity, but rather the triumph of a continuous ideal of authentic becoming. Plato's use of literary structure, that is, his effective use of complex dialogues, illustrates this very act. Plato's work is not a passive jumble of exposition bound together and shelved in time but rather an active discourse engaging readers to confront authenticity in their individual lives. The art of measurement and the power of appearance are real forces that require balancing. Both the white and dark horses of the internal structure of power and the web of interrelated power external to the self require the struggle of authentic becoming. In choosing death as an authentic action, Socrates' choice to refuse to submit to the Athenian power structure exemplifies the struggle for an authentic mode of being.

There remains much ambiguity as to the definition of authenticity as a theme of continental philosophy, particularly whether or not it is a normative ethic. Despite the dissention among the existentialists, authenticity is an important individual ideal. Perhaps authenticity, through the many attempts at existential definitions, has become, not a mode of being, but just another metaphor. Although they neither set forth a declarative treatise on authenticity nor attempt to define it per se, the Platonic dialogues successfully and comprehensively communicate authenticity as a foundational theme. Socrates and Plato actively represent midwives of authenticity, revealing the ideal through the unfolding of a life of struggle such as we witness in Plato's Socrates. This paper has barely touched the surface of the existential thought or the Platonic representations of authenticity, which invite further inquiry. Despite our determined location in a society, as individuals we have the freedom to struggle toward an authentic mode of becoming.

Notes

[1] Marjorie Greene, "Authenticity: An Existential Virtue," *Ethics: An International Journal of Social, Political, and Legal Philosophy* 62 (July 1952): 266-274.

[2] John Arras, "A Critique of Sartrian Authenticity," *The Personalist* 57 (Spring 1976): 171-179.

[3] Arras 77.

[4] Charles Guignon and Derk Pereboom, introduction, *Existentialism Basic Writings* (Indianapolis, IN: Hackett Publishing Company, Inc., 2001) 4-7.

[5] Edwin Sherman, "Authenticity and Diversity: A Comparative Reading of Charles Taylor and Martin Heidegger," *Dialogue* 44.1 (2005): 148-150.

[6] Sherman 151.

[7] Sherman 149.

[8] Gerry Stahl, "The Jargon of Authenticity: An Introduction to a Marxist Critique of Heidegger," *boundary 2* 3.2 (1975): 489.

[9] See Alexander Nehamas, *Virtues of Authenticity: Essays on Plato and Socrates* (Princeton, NJ: Princeton University Press, 1999) xxxii-xxxv and 365.

[10] Nehamas 62.

[11] Sherman 148-150.

[12] *Plato: Complete Works*, ed. by John M. Cooper (Indianapolis, Indiana: Hackett Publishing Co., 1997) 771-775.

[13] *Plato* 30, 33b.

[14] *Plato* 762, 329a.

[15] From *Concluding Unscientific Postscript* cited in Guignon and Pereboom 92.

[16] Guignon and Pereboom, 90.

[17] Guignon and Pereboom 90.

[18] *Plato* 886, 86b.

[19] Charles Guignon, *On Being Authentic* (London: Routledge, 2004) 13.

[20] *Plato* 982, 336e.

[21] Roger Alan Deacon, *Fabricating Foucault: Rationalising the Management of Individuals* (Milwaukee: Marquette University Press, 2003) 167-169.

[22] *Plato* 983, 3338c.

[23] *Plato* 877, 77b.

[24] *Plato* 785, 356e.

[25] Teodoros Kiros, *Self-Construction and the Formation of Human Values* (London: Greenwood Press, 1998) 110.

[26] *Plato* 27, 29d-e.

[27] Deacon 179-180.

[28] *Plato* 26, 28a.

[29] *Plato* 29, 32a.

[30] Guignon 69 and 73.

[31] Guignon 147-149.

[32] Sherman 152.

[33] Sherman 146.

Works Cited

Arras, John. "A Critique of Sartrian Authenticity." *The Personalist* 57 (Spring 1976): 171-180.

Deacon, Roger Alan. *Fabricating Foucault: Rationalising the Management of Individuals*. Milwaukee: Marquette University Press, 2003.

Greene, Marjorie. "Authenticity: An Existential Virtue." *Ethics: An International Journal of Social, Political, and Legal Philosophy* 62 (July 1952): 266-274.

Guignon, Charles. *On Being Authentic*. London: Routledge, 2004.

Guignon, Charles and Derk Pereboom. *Existentialism Basic Writings*. Indianapolis, IN: Hackett Publishing Company, Inc., 2001.

Kiros, Teodoros. *Self-Construction and the Formation of Human Values*. London: Greenwood Press, 1998.

Nehamas, Alexander. *Virtues of Authenticity: Essays on Plato and Socrates*. Princeton, NJ: Princeton University Press, 1999.

Plato: Complete Works. Ed. by John M. Cooper. Indianapolis, Indiana: Hackett Publishing Co., 1997.

Sherman, Edwin. "Authenticity and Diversity: A Comparative Reading of Charles Taylor and Martin Heidegger." *Dialogue* 44.1 (2005).

Stahl, Gerry. "The Jargon of Authenticity: An Introduction to a Marxist Critique of Heidegger." *boundary 2* 3.2 (1975).

Hegel on Community and Conflict

Nathan Andersen, *Eckerd College*

If discussions and debates of the last few decades on the politics of multiculturalism have established anything conclusive, it is that cross-cultural understandings and multi-cultural communities are never pre-given and can never be taken for granted.[1] Moreover, if it were not already obvious from the experience of living in a contemporary world, such discussions have made clear that the problem does not only affect clear cultural borders; every affirmation of solidarity among individuals, every claim to belong to the "same community" or even the same group within a larger community is in fact open to challenge on a number of grounds. We now know that we can never assume in advance that we are actually communicating or that we stand together on common grounds. To put the point another way: "we" can never assume that there is a "we" to assume or that does the assuming. How, then, should we now think the possibility and nature of community? What I want to show is that we can find resources for thinking an answer to this question in what is perhaps a surprising place: Hegel's *Phenomenology of Spirit*. This will be surprising to those who have accepted what has come down as a standard criticism of Hegel: that he takes for granted a readership that shares his metaphysical presuppositions, to whom he refers when as narrator he adopts the royal "we." It should not be surprising, however, for those who know that one of the central problems of the text as a whole is to conceive adequately the character of this "we."

That the universality of the "we" – of Spirit considered as the community that has come to understand itself – is never pre-given in a sense that would allow it to subsume differences between individuals or between cultures, is demonstrated perhaps most powerfully by the culminating analysis of Hegel's chapter on "Spirit" in the *Phenomenology*. There he examines the phenomenon of conscientious action, action rooted in the agent's subjective understanding of the moral demands of her particular situation. I think it does not presuppose too much, in this case, to say that *we* are or at least take ourselves to be for the most part, conscientious agents. This does not presuppose too much, because all it says in the end is that we cannot with full justification presuppose anything else about anybody else. Insofar as we are conscientious, we insist that all action be judged, not according to an external standard or universal principle, but only according to the demands of the situation of action as they are experienced by the agent.

The problem with conscience, however, is that when voiced by the agent herself this insistence calls upon her the judgment that she has violated her own criterion, insofar as she expects

others to act in accordance with a general principle. I, for example, expect others to judge my actions, not according to *their* unique perspective on these actions, not on the basis of their own conscience, but solely in accordance with *my* claim to have acted conscientiously. It is the apparently extreme subjectivity of this demand that at first glance strikes one as being at odds with the values of community. Indeed it appears to preclude the possibility of the mutual understanding and communication that form the basis for genuine community. It is for this reason that Hegel's analysis of the community spirit that emerges out of the conflict of conscience is a special one, with profound bearing upon the question of the possibility of community in a multicultural world.

The conflict of conscience, as we will see, motivates a resolution that does not so much dissolve the conflict as reveal the both inescapable and productive dimension of contrasting self-definitions. The conscientious agent comes to recognize that the *truth* of the situation in which she has an individual obligation to act is not comprised merely of the circumstances of action as they appear to her all at once. Rather, the truth of this situation is what dynamically unfolds as she risks action on the basis of a limited perspective and then finds herself obliged to reformulate that perspective as a result of discovering the compelling character of the judgments of others that this action calls down upon itself.

Hegel's analysis of conscientious activity, and the forms of social life that are set up around this activity, is quite complex, drawing on a fascinating wealth of literature from ancient Greek tragedy and the dialogues of Plato through Protestant and Pietistic writings to the work of contemporaries such as Jacobi, Goethe and Novalis. Much has been written on the literary background of Hegel's work on conscience.[2] For the purposes of the present investigation into the nature and possibility of community, his analysis can be broken into three basic stages, corresponding to three paradigmatic types of relationship that can be set up between conscientious agents. There is, first, the spurious community that is built on mutual assurances of respect without concern for content.[3] The second stage makes clear why this form of community must ultimately fail: this stage makes explicit in social life the conflict that we have already suggested is built into the very idea of conscientious action.[4] The third stage, that Hegel identifies as the stage of forgiveness, indicates important features of a genuine community, one that does not overcome the differences between its individual members, so much as reveal these differences to be both inevitable and productive.[5]

Sharing the same space or belonging to the same economic network of associations, asserting the importance of mutual toleration and respect, are never enough to generate genuine community. All this is, of course, obvious insofar as such assertions hardly serve to stem conflict and aggression, in both its subtle and explicit forms. Even the suppression of conflict or the

overcoming of opposition cannot serve to secure a stable and vital community that truly answers to the needs of a diverse population or group. What Hegel's account suggests is that genuine community, an "I that is we and a we that is I,"[6] can only emerge as we learn to interpret the inevitable conflicts that arise from our differences as legitimate expressions of the true nature of our shared situation.

1. The Mutual Admiration Society: The Community of Conscience Built on Assertions

The basic principle of an ethics rooted in conscience is that an agent can only be responsible to the moral demands of a particular situation insofar as she herself experiences those demands. Since one can never have more than a partial understanding of the total situation wherein one acts, the principle of conscience takes this partially illuminated situation *to be* the *genuine* situation of action, in relationship to which alone the conscientious action can legitimately be evaluated. The ethics of conscience insists, therefore, that all action be judged according to the particular demands of the conscience of the agent alone. The problem with the concept of conscience is that it *appears* to sanction an extreme form of subjectivism as grounds for action.[7] If in the last analysis all we can appeal to is ourselves, and our subjective knowledge of our own situations, this would appear to render community and communication unnecessary, if not impossible. At the same time, even if the particular content of an agent's conscientious grasp of her situation is inaccessible to others – the fact of its uniqueness can at least be understood by others, who can endorse her claim to act conscientiously even if they cannot endorse it of their own right. The community of conscience is, to begin with, rooted in the mutual understanding of the significance of this claim.

The claim of conscience – whose distinctive content is the assertion of the conscientious agent in the face of others that she is indeed acting as conscience demands – aims to give a universal weight to a particular conscientious action. Hegel writes that "in calling itself *conscience*, [the individual agent] calls itself a universal knowing and willing which recognizes and acknowledges others, is the same as them – for they are just this pure self-knowing and willing – and which for that reason is also recognized and acknowledged by *them*."[8] This mutual acknowledgment results in the formation of a community of conscientious agents, a community in which, allegedly, the differences invoked by the principle of conscience are embraced *a priori*. Yet what this community in fact embraces is merely difference in an abstract sense; it accepts a multiplicity of voices, but only to the extent that they are all saying essentially the same thing. It is the community composed of self-proclaimed conscientious agents, and built upon the "mutual assurance of their conscientiousness, good intentions, the rejoicing over mutual purity, and the refreshing of

themselves in the glory of knowing and uttering, of cherishing and fostering, such an excellent state of affairs."[9]

On Hegel's account, genuine universality, the true community of conscience, is not what is achieved in this "mutual admiration society."[10] The *genuine* community of conscience is forged only out of the situation wherein conscientious agents are *unable* to rejoice in the purity of one another, but find themselves in fact engaged in a *conscientious* conflict. It is this situation that forces a recognition of a unity that emerges only in and through difference, and a discovery of the true character of their differences precisely in that unity.

Likewise and in general, true community requires more than merely the mutual admiration of different individuals or groups. To the extent that in the modern world we embrace the idea that for the most part people will act in the ways they deem best, given their background and values, to this extent we should be committed to a general acceptance of others. This general acceptance, however, does not by itself establish community. For it fails to consider the real power of the differences between the individuals and groups that share the same space.[11] Unless these differences are heeded, and unless careful thought and planning is given to the development of social institutions that respond to these differences, they can become the seeds of destruction for the anticipated community. I will now turn to an investigation of the differences that characterize the community of conscience, and then to Hegel's account of the specific manner in which these differences can be addressed in the interests of a genuine community.

2. Conscientious Conflict: The Agent and the Judge

There are, of course, numerous ways in which conscientious agents might come into conflict with respect to a particular content of action. *My* conscience need not and often will not dictate that I be satisfied with the way *you* find *yourself* compelled to act in a determinate situation. Insofar as each of us is committed to acknowledging the authority of conscience, however, we still maintain a certain kind of communal bond. We may fight, and I may decide that it is my obligation to constrain your activity. Yet to the extent I take seriously the principle of conscience, then even as I do so, I cannot help but respect your willingness to put your life on the line in following your conscience. Insofar as I know you to be adhering to what you take to be your duty, we share a common bond.

The real or essential conflict of conscience, however, does not, on Hegel's account, arise as a disagreement about some particular course of action. Rather, Hegel argues, this conflict arises as a result of the fact that there is a tension or ambiguity in the principle of conscience itself. The claim,

on the part of the conscientious agent, that she is acting in accordance with the dictates of her own conscience, can as a result of this ambiguity be taken up in a sense quite different from the one in which it is intended by the conscientious agent. This ambiguity in the utterance of conscience is at the heart of the conflict out of which the true situation of conscience comes to light, according to Hegel.

The problem with the utterance of conscience, as Hegel puts it, is "that the two moments constituting this consciousness [of the agent], the self and the in-itself, are held to be unequal in value within it, a disparity in which they are so determined that the certainty of itself [i.e. the conviction of the conscientious agent that she is acting as she must] is the essential being in the face of the in-itself or the universal [i.e. the acceptance of this utterance by the community], which counts only as a moment."[12] To proclaim the conscientiousness of one's actions is to know oneself as a singular being whose actions carry out a lived experience of necessity that may not be accepted by others, at the same time as it is to insist that these actions be respected by these others. What *matters* for the conscientious individual is the fact that her claim to conscientiousness be genuine and that she recognize herself in this claim; to the extent that she *does* she expects that others will recognize this as well. The aspect of universality, or of universal acceptance, however, is taken for granted. The agent does not, in other words, take this judgment on the part of others to *matter* to the character of her action *as* conscientious. For the agent, her action *is* conscientious, and hence justified, regardless of what others think, and yet she nevertheless expects others to recognize automatically her right to act on the basis of her conscience.

For the judging consciousness, by contrast, it is the judgment *alone* that matters when considering the utterance. Even when the judge takes for granted the *feeling* of conscientiousness that the agent proclaims in the utterance, even when the judge accepts that the agent thinks of herself as conscientious, there still remains to her the task of determining whether the utterance itself is a conscientious act. The judge *knows* the utterance to be justified through the principle of conscience only to the extent that she judges it to be such; yet since the utterance of the agent poses this judgment as a foregone conclusion, as something that need not in fact take place, the judge must reject the utterance as a failure to live up to the demands of the principle of conscience.

The *sacred*, inviolable character of claims to conscience is, from the standpoint of the judge, taken *lightly* by the conscientious agent. From the standpoint of the judge, who knows that she is *not* in a position to judge whether or not the agent in fact operates on an undeniable conviction, the assertion that she *does* appears as a contingent strategy for acceptance, one that indicates her not really to be in earnest about being accepted. She is, in that sense, according to the judge, a hypocrite.[13]

To make this clearer, I will turn briefly to one of the literary sources that Hegel could have drawn upon in his development of the account of conscience; whether or not he had actually read this piece at this point in his career, he was definitely familiar with several similar literary sources, among them writings by the same author. We find a very good example of the kind of conflict Hegel has in mind in the case of Edward Allwill, as judged by Sylli and Lucy and documented in Jacobi's fictional novel entitled *Edward Allwill's Collection of Letters*. The young Edward insists repeatedly that in his actions he only follows his conscience and his heart, even if the actual paths he follows may appear erratic and misguided to others. In fact, however, no one ever denies his good intentions, or argues that his claims to conscientiousness are not sincerely uttered, at least in light of his feelings of the moment. Sylli, whose brother-in-law has taken Edward under his wing, writes:

> The unruly fellow may well be a worthy young man, and one who regularly means to do more good to others than to himself: but that only makes him all the more dangerous; that's what gives him the open, innocent countenance, against which no counsel holds, so that one reaches out one's hand to him from afar, entwines one's lot with his, and makes company with him. Only later does one become aware how uncertain the ways he travels are, how foolhardy he is in action, how cheaply he offers his skin for sale, and hence the skin of his associates as well.[14]

What is reprehensible in Edward, according to Sylli, is not that he acts as he feels to be right in the moment, but that he fails to take seriously the impact that his very expressions of conviction have upon those others who accept them.

Lucy, who is in fact the deceived one that Sylli describes here, who had fallen in love with Edward only to discover later that his heart and conscience were fickle, likewise claims that what is wrong with Edward is not his wholehearted immersion into his various projects, but his failure to consider the effects of these projects upon others. After reminding him of how he had once been seriously troubled by the fact that others took him to be frivolous, and therefore did not trust him, she writes: "How great, how lovely! At that time, how close my Edward to the *best* of his species!"[15] Now, however, on Lucy's account, since he no longer cares for what others think of his actions, as long as he feels himself to be right, he has cast away the *innocence* of the soul that allows the reactions of others to trouble his conscience. From a "dear, sincere – royal youth" he has become "alas, so lowly degraded to … an anxious, squinting sophist!"[16] His sophistry, from the point of view of Lucy, consists in the fact that he makes use of a true principle, the principle of conscience, not in

order to seek out truth or become better, but rather to maintain his feeling of self-complacency in the face of the criticisms of others.

From Edward's standpoint, by contrast, Lucy herself has become the insensitive judge who in her insistence upon the respect for principles has lost the ability to feel, and to follow her conscience as it is expressed in the immediate inclinations of the heart. According to Edward, it is *she* who treats the genuine convictions of others as if they were a matter of little importance when they come up against her own steadfastly adopted principles.

In general, the conscientious agent will consider the judgment that opposes her actions to be itself one-sided, and to rest upon a particular and idiosyncratic way of taking up the principle of conscience in the same way that this judgment proclaims to be true of the agent. The judge does violence to the expectations of the agent, just as much as this judge considers the agent to show disregard for the expectations of others. For this reason, however, it is precisely in her confrontation with the judge that the conscientious agent is able to see herself. The judge takes her own position to be just as obvious as the agent takes hers to be.

3. The Possibility of Reconciliation through Forgiveness

The agent who is troubled by this conflict, and does not, like Edward Allwill, merely brush it off as the result of too much seriousness on the part of the judge, is enabled thereby to learn two important lessons; it appears, moreover, to begin with that learning these lessons will help her resolve the conflict. On the one hand, the condemnation by the judge makes the agent aware that her own way of taking up the principle of conscience is not universally acceptable. It is *not* enough merely for her to be convinced that she does right. On the other hand, because she herself does not interpret the principle in the same way as does the judge, she knows as well that the judge's position also lacks universal acceptability. As a result of the conflict, she has found herself compelled to see the *truth* of the judge's condemnation, and now knows that not only is her course of action controversial but even her *claim* to be acting conscientiously is open to multiple interpretations. For that reason, she can no longer anticipate that others will automatically accept her claims to be conscientious. She recognizes the possibility of failure to communicate as inevitable, and characteristic of her shared situation with others.

Having recognized this, however, she expects that the mutual acceptance of failure will produce a kind of reconciliation. The next stage in Hegel's analysis of conscientious conflict is the confession of the agent that she was wrong to expect immediate acceptance. While she cannot accept the judge's response to her actions, she nevertheless comes to *respect* this response, as a

singular expression of the judge's own conscience. As a result, she is able to admit that she understands why her expression of conscience was unacceptable. She is willing to confess that when she acted, though her choices seemed right at the time, yet her expectation of acceptance did not take into account the conscience of the judge.

To the extent, however, that she expects the judge to accept this confession and admit also that her judgment was one-sided, she merely repeats her error at a higher level. Her confession merely indicates that she has so far *continued* to interpret the judge's position solely in light of her own, and on her own terms. From her perspective, as we saw, what mattered was to act in accordance with the dictates of her conscience, and she expected that *all* would recognize and sanction such a course of action. Having recognized and acknowledged that the judge is also acting as she feels she must in light of her own best understanding of the situation they share, she expects that her confession will be automatically acknowledged and accepted by the judge, in the same way that she had previously expected her claims to conscience to find an immediately accepting audience.

Reconciliation, or forgiveness, can only come, on Hegel's account, when both agent and judge come to acknowledge not merely that conscientious conflict is possible – so that the recognition of the validity of one's claims can never be automatically expected – but that the very character of the claim to conscience calls forth such conflict. Each must come to recognize, in the opposed claims of the other, the experience of necessity that animates her own claims. Moreover, each must come to recognize the essential bearing of opposed claims upon her own conscientious character. As Hegel puts it, "just as the former [agent] has to surrender the one-sided, unacknowledged existence of its particular being-for-self, so too must this other set aside its one-sided, unacknowledged judgment."[17] The agent has recognized the demands of her own conscience to *include* respect for the demands of the other, and therefore no longer takes the *recognition* of her claims to conscience by this other as inessential to her own identity as a conscientious agent. Likewise, the judge must come to recognize that what is essential is not the legitimacy of her judgment *per se*, but *that* the agent herself come to *recognize* the legitimacy of her judgment, and thereby come to judge herself.

Forgiveness, for Hegel, is not merely a matter of accepting and tolerating the finitude and partiality of the stances adopted by others. Forgiveness does, however, come to recognize the futility of a merely *external* criticism. It involves the acceptance of the other in her uniqueness, as a self-determining being who is not in principle confined to the particular, finite stance she appears to adopt, but who will nevertheless accept and respond only to criticism that she can understand on her own terms. At the same time, the forgiving conscience comes to acknowledge that the criticisms of her, made by others, cannot themselves be merely external, and recognizes the demand to come

to comprehend as necessary the experience of her own shared situation with others that called forth precisely their concerns. Forgiveness, in Hegel's sense, is thus not something that one can *decide* whether or not to extend to the other; to forgive is rather to have learned to respond to and acknowledge the other as genuinely conscientious, *both* in the sense that she responds to her own situation as she knows best and in the sense that her own appraisal of a shared situation cannot be ignored or taken lightly.

Conscientious forgiveness identifies as *genuine* community only that which emerges out of particular situations wherein individuals and groups discover themselves and the character of the situation they share with others and with other groups in and through their acknowledgement of the compelling character of the other's affirmation of difference. The community that is achieved between them is not a community that exists prior to their opposition, a shared identity that they come to realize they had already possessed. It is not built upon a tacit contract. It is not even the community that comes by way of affirming differences in advance, and that for that reason overlooks the formative role of differences in establishing identity. It is, rather, the unity that emerges through the dynamic interaction and opposition of distinct individuals, who refuse to be assimilated by each other and yet, because they recognize this refusal in the other as well, discover who and how they are by affirming and identifying with the other in their differences. They discover the uniqueness of their own experienced demands precisely insofar as they open themselves up to a discovery and awareness of the demands that animate the actions of others. This self-discovery, moreover, is just as much a matter of self-formation or self-creation, insofar as it involves the active reformulation of the boundaries between each individual and its others, that takes place as each comes to comprehend ever more adequately the necessity that motivates the activities of these others – expressed explicitly by conscientious agents – to be a necessity according to which she is herself bound. Each must learn to accommodate her own conception of self, and the actions that correspond to that conception, to the realities of a situation defined by competing conceptions.

4. Community as Emerging from the Shared Recognition of the Reality of Difference

We become the (spurious) "we" that we are, in the first place, through our failures to consider the differences that divide us. Such failures can be productive, however, insofar as they establish points of conflict that enable such differences to come to expression. Genuine community does not resolve or overcome these differences, and cannot render them innocuous or irrelevant. Rather, it is only by thinking and acting in ways that render such differences explicit that we can avoid collapse into the spurious community whose shared identity exists in name alone. The results

of this process are always particular, in the sense that the result is always the formation of a determinate shape of community in which uniquely situated individuals have to some degree come to terms with or even defined themselves in relation to their differences. To make such differences explicit involves more than merely embracing what is innocuous or interesting, as in the case of the mutual admiration society we examined above.

Community is never pre-given, and yet the activities, institutions and thoughts through which particular individuals or communities fail to appreciate the distinctive characteristics of others are activities, institutions and thoughts that *do* constitute these individuals and communities in relation to the others they fail to consider. Community can only be created and established by way of the specific encounters wherein opposed individuals and groups reveal to themselves and to each other the nature and implications of their own expectations, precisely as these expectations call forth opposition or resistance from others who are differently situated. Only as divisions, stratifications, and mutual oppositions become explicit can they begin to be addressed. At the same time, it is only as these divisions become explicit for the members of a divided community that the community as a whole can develop a shared sense of what kind of community they are. As differences are made explicit, and with the shared acknowledgement that these differences are real, there can emerge a collective sense that these differences constitute a problem for all, a problem that all are implicated by and all must deal with. Assurances of mutual respect and toleration are empty insofar as they fail to speak to the reality we share, a reality whose tensions cannot be spelled out and forgiven in advance. It is only through the interpretation of the significance of conflicts as they emerge that we develop a collective sensibility of who we really are and what is at stake in our being together. It is this collective sensibility that in forgiveness can abandon futile hopes that our situation should be something other than it in fact is, and only thereby constitutes the real beginnings of genuine community that can learn to develop meaningful changes.

It is not, then, according to Hegel, just that opposed individuals or groups come better to *understand* themselves by relating to opposition. Rather, they *become* themselves only as each begins to acquire *through* the other a genuine sense for the unique character of her own situation. Genuine community is forged only as the thinking, activities, and institutions that divide us as individuals and cultures are taken seriously, which means recognized as mutually defining insofar as they establish the conditions under which alone we can interact fruitfully. Denying difference, by asserting and operating as if at bottom we are all the same and ought to be able to act out the dictates of our own conscience, can in the short run only generate implausible and untruthful optimism and in the long run lead to the kind of radical misunderstandings that result in injustice, violence, and terror.

Notes

[1] See, for example, bell hooks, *Feminist Theory: From Margin to Center* or Charles Taylor, *Multiculturalism and the Politics of Recognition*.

[2] See, for example, Hyppolite's chapter on conscience in his *Genesis and Structure of Hegel's* Phenomenology of Spirit and Sax's "Active Individuality and the Language of Confession: The Figure of the Beautiful Soul in the *Lehrjahre* and the *Phenomenologie*."

[3] Hegel writes of this community that "the spirit and substance of their association are thus the mutual assurance of their conscientiousness, good intentions, the rejoicing over this mutual purity, and the refreshing of themselves in the glory of knowing and uttering, of cherishing and fostering, such an excellent state of affairs," M656, W/C 431. References to Hegel's *Phenomenology of Spirit* throughout this essay will be to the numbered paragraphs in the translation by A.V. Miller, indicated by an M followed by the paragraph number. These references will be followed in each case by references to the German edition of the *Phänomenologie des Geistes*, indicated by W/C and then the page number.

[4] M659-68, W/C 433-40

[5] M669-71, W/C 440-42. My reading of Hegel's notion of conscience is indebted to the excellent studies of Jay Bernstein. See his "Confession and Forgiveness: Hegel's Poetics of Action" and "Conscience and Transgression: The Exemplarity of Tragic Action."

[6] M177, W/C 127

[7] Hegel appears to criticize conscience on just these grounds in the *Philosophy of Right* (in section 3 of Part Two, entitled "The Good and Conscience") in *Elements of the Philosophy of Right*, arguing that the individual must come to acknowledge duties to the institutions that sustain it. Still, to the extent that such acknowledgement takes place, it remains the responsibility of the individual to make sense of what such duties entail. To that extent, the principle of conscience is not rejected in the *Philosophy of Right* but merely complicated in ways similar to the developments of conscience in the *Phenomenology* as they are described here. That is to say, the individual who embraces the ideal of conscience must discover that the significance of his or her own situation and the obligations it entails is not what appears all at once, but appears only as these actions solicit a response from the others and organizations they impact.

[8] M654, W/C 430.

[9] M656, W/C 431.

[10] As it is described by Judith Shklar in *Freedom and Independence: A Study of Political Ideas of Hegel's Phenomenology of Mind* 191.

[11] A similar point can be made against social contract theories in general, insofar as they refer to a pre-established or tacit commitment to living together under conditions that make this possible while failing to acknowledge that the character of these conditions is never simply given but emerges, historically, in an irreducibly particular fashion.

[12] M660, W/C 434.

[13] According to the judge, as Hegel explains, one cannot even consider the agent's employment of the principle of conscience as a sign of respect for its sanctity, for "the fact that [she] uses what is an essence as a *being-for-another* implies rather [her] own contempt for that essence, and the exposure to everyone of its own lack of any substantial being. For what lets itself be used as an external instrument shows itself to be a thing which possesses no importance of its own," M661, W/C 435.

[14] *The Main Philosophical Writings and the Novel Allwill* 428.

[15] *The Main Philosophical Writings and the Novel Allwill* 475.

[16] *The Main Philosophical Writings and the Novel Allwill* 475.

[17] M669, W/C 440.

Works Cited

Bernstein, Jay. "Confession and Forgiveness: Hegel's Poetics of Action," in *Beyond Representation: Philosophy and Poetic Imagination*. Ed. Richard Eldredge. New York: Cambridge UP, 1996.

Bernstein, Jay. "Conscience and Transgression: The Exemplarity of Tragic Action," in *Hegel's Phenomenology of Spirit: A Reappraisal*. Ed. G.K. Browning. London: Kluwer, 1997.

Hegel, G.W.F. *Elements of the Philosophy of Right*. Trans. H.B. Nisbet. New York: Cambridge UP, 1991.

Hegel, G.W.F. *Hegel's Phenomenology of Spirit*. Trans. A.V. Miller. Oxford UP, 1977.

Hegel, G.W.F. *Phänomenologie des Geistes*. Hrsg. von Hans-Friedrich Wessels u. Heinrich Clairmont. Felix Meiner Verlag GmbH, Hamburg, 1988.

hooks, bell. *Feminist Theory: From Margin to Center*. Cambridge, MA: South End Pr, 2000.

Hyppolite, Jean. *Genesis and Structure of Hegel's Phenomenology of Spirit*. Trans. Samuel Cherniak and John Heckman. Evanston: Northwestern UP, 1974.

Jacobi, Friedrich Heinrich. *Allwill*, in *The Main Philosophical Writings and the Novel Allwill*. Trans. by George di Giovanni. Buffalo, NY: McGill-Queen's UP, 1994.

Sax, Benjamin C. "Active Individuality and the Language of Confession: The Figure of the Beautiful Soul in Lehrjahre and the Phenomenologie." *Journal of the History of Philosophy* 21:4 (1983): 437-66.

Shklar, Judith. *Freedom and Independence: A Study of the Political Ideas of Hegel's Phenomenology of Mind*. New York: Cambridge UP, 1976.

Taylor, Charles. *Multiculturalism and the Politics of Recognition*. Princeton, NJ: Princeton UP, 1994.

The Electoral College and Democratic Equality

Joseph Grcic, *Indiana State University*

The US constitution specifies that the president and vice president of the United States are not to be chosen directly by the people but by what has come to be called the electoral college. The thesis of this essay is that the electoral college should be abolished because it lacks a rational justification and contradicts the basic principles of democratic society as articulated by John Rawls.[1]

The electoral college is constituted by electors chosen by each state and the District of Columbia. The number of electors is equal to the number of senators and representatives the state has in Congress plus the electors assigned to Washington, D.C. There is only one constitutional restriction as to the identity of the electors, "No Senator or Representative or person holding an office of trust or profit under the United States shall be appointed an elector."[2] Further qualification, if any, for being an elector and their selection are determined by state legislatures.

The constitution does not require that electors vote as the majority or plurality in their state voted, but some states do have such a requirement or something similar. In Ray V. Blair the US Supreme Court determined states may, but need not, require electors to vote according to popular vote, but there is no clarity about what would happen if some elector chose to ignore the state law.[3] In fact, it seems clear that the framers of the constitution expected the electors to vote as free agents according to their own judgment and conscience which contradicts the court's view in Ray.[4] When the electoral college meets, about six weeks after the election, a majority of 270 electoral votes out of 538 are needed to win the presidency. Failing this, the election is thrown into the House where each state has one vote.

1. Rationale

The justification for creating the electoral college included two main reasons offered by the framers of the constitution. One argument seems to have been based on federalism. Federalism was understood as a sharing of power between the central government and the states and intended to limit the federal government and so help protect liberty. The rural states were concerned that they would be less powerful than the urban and more populous states if the president and vice president were elected directly by the populace. To this end some supported the electoral college as a compromise to appease the small and slave holding states who felt a direct election of the president would place them in a weakened position.[5] The underlying idea seems to have been that the winner

of the presidential race should have wide appeal geographically thus promoting the stability and unity of the nation.

A second rationale for the college was to be found in the thinking of individuals such as Alexander Hamilton who were suspicious of mass democracy. He and other members of the economic elite saw the potential for mob rule by an uneducated, widely dispersed, and morally limited populace who may threaten private property.[6] Individuals such as Hamilton saw the electoral college as consisting of mature, affluent wise white men who had the best interests of the country (and their own) when they deliberated.

2. Critique

The argument against the electoral college presented here has two parts, one consisting of an analysis of the original justification of the institution and one based on John Rawls' political theory, but whose validity are generally acknowledged among many political theorists.

First, the idea of preserving federalism and widespread geographic support for the president by means of the electoral college is dubious. Federalism is not well defined by the constitution, but it is clear it can and has been institutionalized in ways not necessarily incorporating the electoral college. For example, the sharing of power between the federal government and the states has been implemented by, among other structures (discussed below), the bicameral system which gives each state two senators regardless of population. Moreover, geographic diversity is a lesser value in contemporary society which is more economically diverse and mobile. Finally, one must balance the value of geographic diversity, to the degree that there is such a value, with the values compromised by such a value, such as political equality and democratic rule (argued below). The very fact that slavery played a role in the creation of the electoral college should be an indicator of the nature of the institution.[7]

Second, contrary to Hamilton's assumptions and fears about "mob rule," the population is far more educated than at the time of the writing of the constitution. More importantly, there are many other institutional structures in place that are designed to prevent "mob rule," though the exact meaning of this phrase is not clear. Federalism, separation of powers, checks and balances, judicial review, independence of the judiciary, representative not direct rule, the bicameral legislature (where the House members, elected for two years, are said to be more responsive to current popular opinion and the Senate elected for six years is supposedly more immune to the ephemeral moods of the populace), the party system, and the free press all serve to make mob rule less likely. In addition, various contemporary social welfare programs and a regulated and more stable economy make less likely the possibility of a desperate mob intent on violating the property of the elite. Finally,

democratic theory, as will be explored more fully below, is not based on whether the populace is educated or wise but on the equal moral rights of persons.

Third, the other main justification for the college, that it would be a deliberative body of wise or at least better informed individuals who would select the president and vice president in a rational manner, has not been implemented. The electors are not chosen because of their wisdom, even if that were possible, since there is no generally accepted legal definition of "wisdom." If wisdom is interpreted to include a kind of spiritual or religious dimension, then it would violate the establishment clause, or the separation of church and state, as specified in the First Amendment. In fact, the electors are chosen usually for their loyalty to their political party.

In Ray V. Blair mentioned above, the court ruled in favor of the claim that electors are not necessarily free agents but can be required to vote according to party loyalty or popular vote. The court's ruling would seem to imply that wisdom is not a relevant criterion.

A further problem which exacerbates the moral and political problems associated with the electoral college is the problem of unfaithful electors. Unfaithful electors, at least seventeen so far, are those who did not cast their vote as they promised or as the popular vote majority dictated. At times these electors acted as they did in exchange for personal favors from those elected.[8]

3. Rawls

The problematic and anomalous aspects of the electoral college are made more perspicuous by comparison with John Rawls' theory of justice.[9] Rawls considers himself to be working in the tradition of the Declaration of Independence, the basic principles of the US constitution and the social contract theory of Locke, Rousseau and Kant who understood a democratic system to be, essentially, a society where citizens are defined as "free and equal."[10] According to Rawls, a society based on the principles of democratic equality would be based on his two basic principles of justice. The first principle states "each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others." And the second principle holds: "social and economic inequalities are to be arranged so that they are both a) reasonable expected to be to everyone's advantage, and b) attach to positions and offices open to all.[11]

The electoral college is a violation of Rawls' first principle in two senses. First, the college denies the general population the right to select the president and vice president directly and as such jeopardizes the basic structure as defined by the two principles. The very existence of the college contradicts democratic principles of equality since it implies that some persons are more qualified than others to elect the president and vice president. Moreover, the electors are not necessarily elected by the citizens of each state and the votes of the electoral college as a whole have not always, nor are they legally required to, correspond to the winner of the national popular vote.

Furthermore, because the college incorporates the distribution of political power which allocates two senators to each state regardless of population, even if electors vote consistently with the popular vote, the result would not guarantee that the winner of the popular vote would win the presidency.[12] Because of this fact, the rule of one person, one vote is violated since the vote of individuals in small states literally is more powerful than the vote of a voter in a large state. Thus, for example, in the 1988 election, one electoral vote in Alaska for every 66,705 voters while in New York, there was one elector for every 178,604 votes.[13]

The two principles, according to Rawls, would form the foundation of a well-ordered society. This is a society where there is a conception of justice which the citizens know and accept and the basic structure of society satisfies these principles in general. In addition, citizens in a well-ordered society have a sense of justice so that they apply these principles and act accordingly for the most part.[14] Finally, a Rawlsian well-ordered society is one where the basic structure promotes the virtues of fairness and trust among the citizenry.

The electoral college contradicts the idea of a well-ordered society because the structure of the electoral college is under-determined and functions at the discretion or whimsy of many individuals at many levels. At the level of selecting electors, there are variations and inconsistencies and once electors are chosen their decisions are essentially undefined especially given the constitutional uncertainties mentioned above. If the vote is thrown into the House, there are no restrictions as to how the representatives may vote (other than one vote per state) which introduces further uncertainty especially since the House vote is subject to bartering, and possible bribery, further removing it from the sovereignty of the people. Given these facts it is clear that the electoral college is not a sufficiently defined element of the basic structure and thus contaminates the basic structure with vagueness which undermines its function of defining the basic rights and duties of the populace.

The electoral college also conflicts with the Rawlsian requirement that the basic institutions encourage the virtues implicit in the basic structure such as trust. Since the electoral college can, except in the cases where limited by state law, act in virtually any way it chooses, it conflicts with the virtue of trust since trust cannot exist without sufficient structure of mutual respect and reciprocity or in a deceptive framework. The current presidential election process is also deceptive in that the popular vote takes place as if the citizens choose the president and vice president directly which contradicts purpose of the electoral college.

4. Stable Society/Public Reason

For Rawls, a well-ordered society is also a stable society if it meets certain conditions. The basic idea of stability, as Rawls understands it, is that there is a coherence between the political system and human nature and human psychology.[15] The basic principle here is the well established ought implies can; that no person can be asked to do what is impossible for humans to do, either physically or psychologically.

The specific requirements of stability Rawls articulates with three points. First, a stable constitutional society is one which must "fix, once and for all, the basic rights and liberties, and to assign them a special priority."[16] Here, for example, Rawls believes that utility fails the test of stability since it does not guarantee the basic principles of justice and as such creates too much uncertainty in the social system. Specifically, Rawls claims that the principle of utility could violate the ought implies can principle by requiring persons to do more than is psychologically possible.[17]

Second, in a stable system, the basic structure of the society is the basis of "public reason" and the basic institutions encourage the virtues of "public life" and "fairness." By fairness Rawls means that institutions satisfy the two principles of justice and that one has "voluntarily accepted the benefits of the arrangement or taken advantage of the opportunities it offers to further one's interests."[18] And third, a stable society is one where "its basic institutions should encourage the cooperative virtues of political life." These virtues are those of "reasonableness and a sense of fairness, and of a spirit of compromise and a readiness to meet others halfway."[19]

The electoral college is clearly incompatible with Rawls' idea of stable society. Specifically, it conflicts with the first requirement that the basic rights be "fixed." The electoral college is, as argued above, fundamentally under-defined in that the rights specified in the basic structure of the constitution are open to violation.

The electoral college also conflicts with the second requirement of a stable society in two ways. First, it conflicts with the idea of what Rawls terms "public reason."[20] Public reason is an ideal of democratic citizenship and fairness where citizens agree that political discourse in a just society will be open and public and in terms of the basic rights and duties embodied in the mutually acknowledged principles defining the basic structure. Public reason is public in three ways: (1) it pertains to the reasoning of citizens as free and equal; (2) it is open and public discourse about the general good and social justice; (3) the content of the discourse is given by a public conception of justice and the basic structure.[21]

The content of public reason involves constitutional essentials and basic justice to be settled by political values alone. The legitimacy of public reason is based on whether or not it uses concepts and principles constitutive of the overlapping consensus. Public reason does not apply to

private deliberations about comprehensive doctrines but only to political and constitutional essentials of the basic structure.

The electoral college is fundamentally a violation of public reason. The college does not respect citizens as free and equal since the average citizen does not directly elect the president nor is there a guaranteed legal requirement that his or her vote actually counts. The college may function independently of the values implicit in public reason and not engage in public debate at all. Public justification of electoral votes cast, whether in conformity to the majority or not, is not required by law. The electoral college functions in a manner not engaging with public reason or based on the content of the basic structure but in private deliberations. Certainly a free and equal people would not agree to the idea of a system of election which operates in an anomalous manner which potentially contradicts with the basic structure.

5. Rule of Law/Legitimacy

The Rawlsian ideal of fairness and a well-ordered and stable society all assume the ideal of the rule of law. The rule of law ideal sees laws as specific norms generated by the basic principles of justice as applied to the social circumstances and conditions in a given domain. According to Rawls, a just legal system must be consistent with the basic principles of the basic structure and conforms to the rule of law if the system of laws is clear, public, complete, and consistent set of norms which specify how members of the community and institutions should behave and function. Rawls adds that the rule of law in its implementation requires that similar cases be treated similarly and that there can be no criminal offense without a corresponding law, no law can violate the 'ought implies can' principle, in addition to some other conditions.[22]

To be sure, the rule of law, as opposed to what is often termed the rule of men, is an ideal abstraction. It is an abstraction since laws are written in a language vague to various degrees which must be interpreted and implemented by imperfect and variously biased persons. Nevertheless, the rule of law sets some limits to these variables and provides somewhat stable parameters which make possible law as a guide to human action and as the basis of social order.

The electoral college clearly violates the ideal of the rule of law. As already indicated, the college is not well defined in structure which raises the question of its coherence with the basic structure. As also indicated, various state laws governing electors in the individual states are variable and potentially inconsistent with each other and the basic structure. In sum, it seems that the main conditions of the rule of law, namely clarity, consistency and completeness, seem to be violated by the electoral college.

The contradictions between the electoral college and the rule of law raises the issue of the legitimacy of the political system. In general, the legitimacy of political power is determined by

whether the political system functions consistently with the moral and legal parameters of the society within which it exists. Rawls understands the legitimacy of political power as existing, "only when it [political power] is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason."[23] One need not subscribe to every nuance of Rawls' idea of legitimacy to see that the electoral college, as already discussed, is deeply flawed and based on an ideology which does not accept persons as free and equal. It is also a commonplace of history that women, Native-Americans, African-Americans, slaves and men without property were excluded from the constitutional convention and its ratification.[24] These fundamental inadequacies in the creation, ratification, and implementation of the constitution allowed the creation of the electoral college, as well as the acceptance of other institutions, such as slavery, which contradict the fundamental ideal of democratic equality.

Finally, the anomalous nature of the electoral college brings forth a dilemma. If the electors vote for the candidate who has received the most votes, they are redundant. If they do not, and, as already indicated, there is no legal guarantee the electors will vote as those who voted for them expected them to, then the college is inconsistent with the ideals of democratic equality and thus contrary to the conditions necessary for a legitimate political structure.

6. Proposal

If the above arguments are sound, then the electoral college should be abolished through a constitutional amendment or various proposals that would generally have the same results.[25] The president and vice president should be elected directly by the free vote of the citizenry. The electoral college is an anomaly within the basic structure and one of the last remaining vestiges of the inegalitarian elements of the constitution.

Notes

I am grateful to the reviewers of this journal for their insightful comments to an earlier draft of this essay.

[1] This should not be taken as assuming Rawls' theory is correct in every aspect, surely no theory is, but only that the Rawls' ideas used here to argue against the electoral college are relevant and plausible in this context.

[2] "Each state shall appoint in such a manner as the Legislature thereof may direct a number of Electors, equal to the whole Number of Senators and Representatives to which the State may be entitled in the Congress" (Article II, Section 1, US Constitution); cf. Max Farrand, *The Framing of the Constitution of the United States* (New Haven: Yale UP, 1913) 77-8, 85.

[3] Ray V. Blair, 343 US 214 (1952); in McPherson V. Blacker (1892) the court seems to have held a different view; see also Kermit Hall, ed., *The Supreme Court of the United States* (New York: Oxford UP, 1962) 881.

[4] George C. Edwards, *Why the Electoral College is Bad for America* (New Haven; Yale UP, 2004) 25-6.

[5] Slaves were to be counted as 2/3 persons was part of the compromise as well.

[6] Alexander Hamilton, *Federalist* #68, James Madison, *Federalist* #10, 39, John Jay *Federalist* #64; Justice Harlan in Williams V. Rhodes 393 U.S. 23 (1968) accepts this view of the framers intentions in creating the electoral college; see also Charles A. Beard, *An Economic Interpretation of the Constitution of the United States* (New York: Macmillan, 1935) 214 and John F. Manley and Kenneth Dolbeare, eds., *The Case Against the Constitution* (London: Sharpe, Inc., 1987) 31-50; for criticism of Beard see Robert *E.* Brown*, Charles Beard and the Constitution* (Princeton: Princeton UP, 1956) esp. 99, 123.

[7] Lucius Wilmerding, Jr., *The Electoral College* (New Brunswick: Rutgers UP, 1958) 11-12.

[8] Six times the winner of the popular vote did not win the majority of electoral votes. Twice the decision for president was thrown into the House of Representative as provided by the 12 amendment. (In 1824 Andrew Jackson won plurality but lost to John Quincy Adams in the House election; in 1876 Samuel J. Tilden won the popular vote but not the presidency; Benjamin Harrison won by electoral votes in 1888 even though Grover Cleveland won the popular vote; in 1960 John F. Kennedy received fewer votes, and more recently in 2000 Gore received more popular votes than President George W. Bush.) Given the current legal situation, there is a probability of one in three that there will be a lack of consistency in the popular and electoral vote if the margin of popular votes is less than 300,000. If the margin is less than 1.5 million, the probability of an inconsistency is one in four. Even if the vote did correspond to the popular vote, due to the allocation of two

senators to each state regardless of the population, the electoral vote does not necessarily mean the candidate with the most popular votes wins.

[9] John Rawls, *A Theory of Justice* (Cambridge: Harvard UP, 1971) 453-62. The use of Rawls' theory is not meant to imply that only his theory is capable of showing the contradictions and moral bankruptcy of the electoral college. Indeed, the ideas Rawls uses, the rule of law, stability, well-ordered society, etc., are widely shared among liberal democratic theorists such as Ronald Dworkin, *Is Democracy Possible Here*? (Princeton: Princeton UP, 2006) and Robert Dahl, *On Democracy* (New Haven: Yale UP, 1998).

[10] Rawls 19.

[11] Rawls 60.

[12] Bruce Ackerman, *Bush* V. *Gore* (New Haven: Yale UP, 2002) 7-12.

[13] Nelson Polsby and Aaron Wildavsky, *Presidential Elections* (New York: Free Press, 1991) 43.

[14] John Rawls, *Justice* a*s Fairness* (Cambridge: Harvard UP, 2001) 8-9.

[15] Rawls, *A Theory of Justice* 398-90, 499; Samuel Freeman, ed., *John Rawls: Collected Papers* (Cambridge: Harvard UP, 1999) 105-6, 177.

[16] Rawls, *A Theory of Justice* 454; John Rawls, *Political Liberalism* (New York: Columbia UP, 1993) 38-40.

[17] Rawls, *Justice as Fairness* 115.

[18] Rawls, *A Theory of Justice* 110-12.

[19] Rawls, *Justice as Fairness* 116

[20] Rawls, *Political Liberalism* 213

[21] Rawls, *Political Liberalism* 223-225; John Rawls, "The Idea of Public Reason Revisited" in Freeman, ed., *Collected Papers*: *John Rawls* 573-575; Charles Larmore, "Public Reason" in Samuel Freeman, ed., *The Cambridge Companion to Rawls* (Cambridge: Cambridge UP, 2003) 368-375.

[22] John Rawls, *A Theory of Justice*, revised ed. (Cambridge: Harvard UP, 1999) 206-11; Rawls acknowledges his sources here include H.L.A. Hart's *The Concept of Law* (Oxford UP 1961) and Lon Fuller's *The Morality of Law* (Yale UP 1964).

[23] Rawls, *Political Liberalism* 137; Freeman, *John Rawls: Collected Papers* 578-9.

[24] Howard Zinn, *A People's History of the United States* (New York: Harper Perennial, 1995) 88-90.

[25] Robert A. Dahl, *How Democratic is the American Constitution?* (New Haven: Yale UP, 2003) 86-7.

Works Cited

Ackerman, Bruce. *Bush* V. *Gore.* New Haven: Yale UP, 2002.

Beard, Charles A. *An Economic Interpretation of the Constitution of the United States.* New York: Macmillan, 1935.

Brown, Robert E. *Charles Beard and the Constitution.* Princeton: Princeton UP, 1956.

Dahl, Robert A. *How Democratic is the American Constitution?* New Haven*:* Yale UP*,* 2003.

Dahl, Robert. *On Democracy.* New Haven: Yale UP, 1998.

Dworkin, Ronald. *Is Democracy Possible Here*? Princeton: Princeton UP, 2006.

Edwards, George C. *Why the Electoral College is Bad for America.* New Haven; Yale UP, 2004.

Farrand, Max. *The Framing of the Constitution of the United States.* New Haven: Yale UP, 1913.

Freeman, Samuel, ed. *The Cambridge Companion to Rawls.* Cambridge: Cambridge UP, 2003.

Freeman, Samuel, ed. *John Rawls: Collected Papers.* Cambridge: Harvard UP, 1999.

Hall, Kermit, ed. *The Supreme Court of the United States.* New York: Oxford UP, 1962.

Manley, John F. and Kenneth Dolbeare, eds. *The Case Against the Constitution.* London: Sharpe, Inc., 1987.

Polsby, Nelson and Aaron Wildavsky. *Presidential Elections.* New York: Free Press, 1991.

Rawls, John. *Justice* a*s Fairness.* Cambridge: Harvard UP, 2001.

Rawls, John. *Political Liberalism.* New York: Columbia UP, 1993.

Rawls, John. *A Theory of Justice*. Cambridge: Harvard UP, 1971. Revised ed. Cambridge: Harvard UP, 1999.

Wilmerding, Lucius, Jr. *The Electoral College*. New Brunswick: Rutgers UP, 1958.

Zinn, Howard. *A People's History of the United States*. New York: Harper Perennial, 1995.

Values as a Political Metaframe

James E. Roper, *Michigan State University*

1. Introduction

According to noted linguist George Lakoff, framing an issue amounts to placing it in a particular context for evaluation.[1] In the summer of 2005, the Michigan Legislature took up a bill to allow a gasoline pipeline to be built in south Lansing. Those who objected to building the pipeline in this location talked about possible environmental damage and danger to the people living along part of the route (people whose incomes are relatively low). When presented with these arguments, an aide to the state senator who introduced the bill said, "This legislation is *about jobs and economic development for the state of Michigan*."[2] In short, she reframed the issue as being "…about jobs and economic development…" The tacit implication is that the legislation is *not about the environment or danger to those who live along the route*. Are there environmental and physical dangers associated with the proposed placement of the pipeline? The senator's aide did not deny this. Instead she placed the pipeline issue in a context favorable to the senator's (and oil company's) desire to have the pipeline placed in the proposed location. A major goal of this paper is to rise above the false choice encouraged by the presentation of this issue.

Since the 2004 Presidential Election, there has been a lot of talk about "moral values." We heard such talk again during the 2006 Congressional election; and it has already surfaced in the early phases of the 2008 Presidential Election. It is claimed that Republicans did well in the Presidential Election of 2004 because of "their commitment to moral values." Lakoff has suggested that Democrats need to "frame" issues in terms of *their* moral values.[3] Lakoff is correct to see "framing" as one key to how people understand politics, and the suggestion that Democrats need to focus attention on *their* values undoubtedly has merit. Indeed, Lakoff goes through a number of important issues and discusses how they might be reframed in more "progressive" ways. For example, some Republicans have spoken of taxes as a "burden"; but Lakoff suggests Democrats should shift the focus to what taxes do for us. The appropriate frame, he suggests, is to think of taxes as (essentially) the "dues" we pay to live in a country that provides freedom, education, and so on to its citizens.[4] In spite of his profoundly important work on framing, however, I believe Lakoff and others are overlooking something very important.

2. Two Different "Political Metaframes"

As a preface to my substantive remarks, I recount the following true story. I once attended a discussion group that focused on "moral values." A close friend, a prominent member of the Psychology Department at Michigan State University, convened this discussion group. This psychologist had designed a questionnaire to determine an individual's moral values. He had listed twelve moral values that could be ranked from 0 to 9—9 indicating the strongest support. In general, a moral value can be stated without using a sentence. Things like national security (or, more generally, safety), love, benevolence, patriotism, and so on are examples of moral values. One of my psychologist friend's moral values was "honesty." I recall asking him, "Honesty what?" I meant that 'honesty' is so vague it might stand for just about anything—especially if one can rank it from 0 to 9. I pointed out that serious work in ethics and social philosophy usually deals with "moral principles." Principles must be stated in sentences; they cannot be embodied in a single word or phrase the way values can. Although moral principles are by no means perfect and have their detractors, they are typically more specific than moral values.[5] Stating positions in terms of moral principles does not constitute an alternative way of *framing* our views; instead, casting our views in terms of the moral principles we espouse constitutes what I call an alternative "*political metaframe"*—a framework far more precise and difficult to manipulate than the vague language of "values." Indeed, there is an important philosophical reason for this.

A political metaframe is an overarching structure encompassing various different ways of framing a political issue. Principles constitute an alternative political metaframe to that of *values. The United States Constitution* is essentially a set of principles. The Constitution's First Amendment does not say: "[The value of] freedom of religion." It says: "Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; …"[6] This is a principle. Values and principles constitute two different "political metaframes"—two different ways to think about the ethical ramifications of political issues. Principles are more specific than values. Unquestionably, principles can be vague and ambiguous; but they are typically far less so than moral values. Hence, values are more likely to open the door to manipulation and deception. It is true that the preamble to the Constitution contains reference to values, but these values are instantiated in the body of the document as a list of principles.

It is not just the fact that principles are less vague and ambiguous than values, however, that makes a "principles" political metaframe superior to a "values" metaframe. *The key difference between values and principles is that the latter sustain logical inferences, while the former do not, and this is crucial not only to their greater precision but also to their usefulness in political dialogue and discussion.* An excellent example of

how this works in practice appeared in an article about the Supreme Court's decision rejecting Guantanamo Bay tribunals, as the following quotation makes clear:

> In a November 2001 decree that Mr. Bush styled a "military order," he had authorized military commissions to try defendants he selected, according to rules he created, for crimes he defined. But Justice John Paul Stevens, in a 73 page opinion for the court, joined by four other justices, went piece-by-piece through the legal theories the president had asserted, finding in each instance that they ran afoul of law and precedent.[7]

In other words, the principles Bush had enunciated in his "decree" ran afoul of principles that constitute our law, including its clarifying precedents. Hence, using principles led to refinement by virtue of logical interaction with other principles; this, in turn, made political dialogue about the issue in question more productive. A values metaframe does not seem amenable to such analysis.[8]

3. A Critical Response: "Values Hierarchies"

Supporters of using a metaframe of values, however, might question this distinction. Referring, for example, to the work of my psychologist friend with his values questionnaire, such values proponents might argue that values can at least be rank ordered into a "values hierarchy."[9] If this is true, surely values can be compared with one another much as principles can; and my contention that we need to replace a "values metaframe" with a "principles metaframe" could lose some, or all, of its appeal.

To answer this charge, consider the nature of a value such as "national security." What might it mean to rank it on a scale of '0' to '9'? Theories of rational choice often speak of "multidimensional" comparisons of different payoffs, or values.[10] Someone who prefers chocolate ice cream to vanilla ice cream in the "abstract" (whatever that really means) may have a very different preference when asked to compare these flavors under different circumstances. For example, suppose the question is not whether one would prefer chocolate or vanilla ice cream in a bowl in isolation from other food one might be eating; suppose rather that the choice is between chocolate and vanilla as a topping on a slice of apple pie. Again, a man or woman may prefer the company of another man or woman in the context of selecting someone to help with a particular computer project, but have a very different preference if the issue is selecting a companion with whom to have dinner. Abstract comparisons of outcomes are typically useless in helping make decisions. Similarly, abstract comparisons of "values" without any reference to specifying principles

are like cars without steering wheels; a rhetorical nudge can send them in virtually any direction. Selecting just one simply begs the question.

4. Framing and Balancing

Let us return to *national security*—a value many embraced after 9/11. Bush strategists insist national security goes hand in glove with greatly diminished civil liberties, and this attitude drove the original U.S.A. Patriot Act that was rushed through Congress in the wake of the events of 9-11.[11] For Bush, national security is joined at the hip with a radical *principle* that says *national security always trumps traditional American civil liberties and the rights these civil liberties entail.* This marriage of explicit value and implicit principle drives the Bush administration's policy agenda of compromising any right that conflicts with national security, even if the conflict is minimal. Examples abound but some of the most grievous involve gutting the Fourth Amendment—for instance, allowing government agents to look at a citizen's library records and bookstore purchases without a judge's order, as previously required by law.[12] Then, of course, there is the NSA wiretapping program, which Bush officials claim does not even need any legal justification beyond Presidential edict.[13] It is *simply assumed* that embracing the value of national security entails radically compromised civil liberties. If you reject this evisceration of rights, Bush people accuse you of rejecting national security—just as they have said Americans who questioned going to war in Iraq were "traitors" and "unpatriotic."[14]

Recall that Bush said, "You are either with us or against us."[15] This appears to be a *fallacy of faulty dilemma.*[16] Someone who says "The coffee is either hot or cold" ignores the vast range of possible temperatures between those two extremes. Yet the Bush administration's embrace of the value of national security does not allow any compromise with civil liberties: You are either for national security or against it; *you are not allowed to balance national security with other values.*

Such balancing is precisely what a political metaframe of principles promotes. Working directly with principles rather than values forces us to consider the logical relations between different, *and usually conflicting*, principles. Bush typically uses the language of values to say something that can be interpreted as what he really means. His "base" understands what he says in terms of one set of principles—those consonant with their radical ideology—while people who do not share the radical views of the current administration interpret Bush's "values talk" in terms of a different set of principles—one more consistent with the recent history of the country and the Constitution as it has been understood for generations. A great deal of psychological research suggests that people hear what they want to hear if they are given the option to do so.[17] This is especially true if the principles

that demarcate what the administration is really saying are quite radical and the *status quo* suggests a more standard interpretation.[18] For the Bush administration to talk about a value like national security without specifying the principles that circumscribe their understanding of this value is to invite, indeed to encourage, such interpretation in terms of principles that sharply diverge from those that delineate the administration's view of the matter.

On the other hand, if we insist that the administration lay out its positions in terms of *principles*, it is much more difficult to invite the unwary to see things through the lens of the–very different—set of principles that is more consistent with recent history. Whatever form the administration's national security principle takes, it must avoid conflict with a wide array of other principles. This requires balancing principles against one another. Those who work within the political metaframe of values are usually able to identify their *ends* without specifying the *principled means* for achieving them. Potential conflicts are still there, but the metaframe of values keeps them below the surface where the conflicts are hidden—until policies are put in place and it is too late.[19]

In speaking about balancing, we must not obscure the very important debate between those who would interpret rights as absolute in the sense that only conflicts with other rights can supersede them and those who interpret rights in terms of utility and "the greater good." At the policy level, it is very tempting to "weigh" things essentially in terms of "utility," which is often equated with money. On the other hand, the Constitution seems to give rights a different status—a position the contemporary philosopher Ronald Dworkin refers to when he says, "Rights are trumps."[20] Perhaps the most important reason for working in terms of a metaframe of principles is that it allows a clear statement of Dworkin's view in terms of the principles used to justify various policy positions. This is a complex matter and we will return to it, but it is important to make clear that balancing does not mean *"weighing" in terms of utility*; rather, it signals adjustment that recognizes that some principles may have priority over such utility weighing—at least in the eyes of many protagonists.

5. An Exception and a More Typical Strategy

There is one blatant exception to the Bush administration's strategy of presenting their positions in terms of values. In stating his position on stem cells, Bush enunciated a specific principle. He said that it is wrong to take lives in the interest of saving lives.[21] This is extremely interesting. It shows precisely why the Bush administration tends to lay out their positions in terms of values rather than principles. Think of the areas of domestic and foreign policy in which the Bush administration has embraced the principle that the end justifies the means—and specifically

that "collateral damage" may be necessary in order to achieve their political goals. Iraq is a prime example. How many Americans have died, and are continuing to die, in this "action" that Bush has repeatedly tried to convince the American public is the front line of the "war on terror"? How many Iraqis? And neither figure includes those seriously injured or the actual monetary drain on this country and all of the things that money could do actually to save people's lives, both in the U.S. and around the world. If Bush is really committed to the principle that it is wrong to take lives in order to protect lives, surely the Iraq war constitutes a serious counterexample.[22] Indeed, there are many other areas in which Bush administration policies seem tied to an "end justifies the means" position.[23]

In fact, Iraq provides examples of the more typical administration strategy: casting positions in terms of values that are proxies for very radical principles and policies. The administration justified going to war by claiming that Saddam Hussein had weapons of mass destruction he would use against the U.S. His willingness to do this was proved by Hussein's supposed complicity in the 9/11 attacks. When these "justifications" were refuted by seemingly incontrovertible evidence, the administration found a new war warrant in bringing the Iraqis "freedom" and "democracy." The administration has repeatedly spoken of dedication to these fundamental *values*, but the *principles* that circumscribe this allegiance have never been explicitly affirmed.

6. Values versus Principles: Facilitating Productive Debate

There are two characteristics any serious political debate must have: *argument development* and *clash*.[24] By "argument development," I mean that the positions of the disputants must change and "develop" in response to the objections brought up by the other side.[25] For this to happen, of course, the arguments of the opposing sides must "clash,"—that is, these arguments must specifically reply to each other, not miss each other like "trains passing in the night." A political metaframe of values is simply not specific enough to guarantee clash and argument development. Consider the pipeline case with which I began this paper. The homeowners support the value of safety—a more general variant of national security. ("The pipeline is near our houses and it might explode.") On the other hand, the Michigan Legislature is promoting the value of "jobs." ("Putting the pipeline in the particular location the oil company has chosen will mean jobs building the pipeline and jobs operating it after it is built.") Consider what happens when we cast the debate in terms of a political metaframe of principles. Homeowners will emphasize principles such as "Poorer members of society should not be treated unfairly by the State's decisions about where to build things" and "The rights of citizens to have their property protected by the State should not be

outweighed by economic considerations except in the most extreme circumstances (such as urban 'blight')." The Legislature will also focus on principles; for example, "Michigan should do everything it can to work with industry to create jobs in the State" and "The State has the right to determine the best place to put things like pipelines in order to promote job growth in Michigan." It would then become possible for the homeowners, who were probably *most* concerned about the impact of the pipeline on the value of their property, and the Legislature, which was concerned to please the oil company and attract jobs to Michigan, to move from the initial *clash* of their arguments to studying the principles that *ground* these arguments. Perhaps the pipeline could be built in such a way that it would be safer for the surrounding community—burying it deeper, using different kinds of valves, and so on. In addition, those living along the route might have been offered appropriate compensation for the risks and inconveniences associated with the pipeline project. Once the matter was cast in terms of principles, it would be more likely that a solution could be found that would satisfy both parties. For example, the Legislature might approve the project with the understanding that being fair to those living along the route and protecting them from possible hazards might be more expensive but would avoid creating the perception that the State did not care about less well off individuals and their property rights. In short, cooperation becomes possible based on the fact that the clashing principles stand in logical relations to one another. This argumentative development allows various compromise solutions to be considered that would be beyond the reach of a values metaframe. Debating the matter in a metaframe of values holds the wrong things "constant" in the argument and turns the dispute into a media frenzy about "safety vs. jobs." The principles of clash and argument development typify the best academic (high school and college) debate and arguably other political disputes. A principles metaframe is consistent with these guidelines; on the other hand, clash and argument development are not likely to emerge in a dispute framed in terms of values alone.

This discussion of the pipeline issue in the context of academic and political debate demonstrates another benefit of using a metaframe of principles: using such a metaframe may provide a counter to the press' mantra that there are always *two sides* to every story—even if the great majority of real experts support one of the sides. Because principles, unlike values, stand in logical relations to one another, it is much more difficult to argue that someone who objects to the NSA wiretapping program Bush claims he has the authority to keep in place is simply a "traitor." This dispute clearly turns on a number of complex principles that are interconnected with one another. The President claims, for example, that Congress gave him authority to run such a program; but the laws of the nation, which are principles, contain provisions that appear to contradict this claim. Someone who disputes the President's word is not obviously a traitor. He or she may, in standing

up to an unjustifiably expanded claim of executive authority, be taking a very courageous stand against unwarranted claims to authority. Even if this turns out not to be the case, such an individual should not, as a matter of course, be labeled a "traitor." The willingness of the press to say, in effect, "Well, that's the Bush administration's position; it deserves to be taken as seriously as any other view" makes some sense if we are working within a realm of values. ("You are for national security or against it.") In a sphere of principles, however, such simple dichotomies immediately and obviously break down.

7. Principles and Policies That Hide Behind Values

When asked why we are in Iraq, some Americans have cited the value of freedom and the Bush administration has capitalized on this by framing its appeals for support around this notion.[26] We can explore what principles and policies define "freedom" for the administration by looking at their policies in venues where this "value" is invoked. If my thesis in this paper is correct, we should find the administration cynically citing its values but making no mention of the principles or corresponding policies that it takes to be implicit in these values. I label this behavior "cynical" because I believe the administration knows full well that many Americans have a very different view of freedom than the one that drives the Bush policy agenda.[27]

Earlier we spoke of the administration's emphasis on the value of "national security," especially in the lead-up to the 2004 presidential election. The administration's value of "freedom" is especially interesting because it represents the other side of the "national security" coin. In promoting national security, the principle that national security always trumps civil liberties, and policies derived from such liberties, appears to be the order of the day. Such policies are enshrined in the U.S.A. Patriot Act, as we have seen above. Since its hasty passage by Congress in the wake of 9/11, civil libertarians from all sides of the political spectrum have repeatedly attacked the U.S.A. Patriot Act.[28] The weakening of rights such as due process, freedom from arbitrary arrest, search, and seizure, and *habeas corpus* imply that the *principle* this administration followed here is similar to that tacitly invoked under the banner of "national security": *whenever there is a conflict between fundamental rights and the government's interest in "combating terrorism"* (or promoting "national security"), *the government must prefer the latter—no matter how small the probability of some sort of attack on the country is or the probable harm from such an attack.*[29] (The recent "revision" of the U.S.A. Patriot Act did little to alter this commitment.[30]) Again, it does not seem to be a matter of *balancing* fundamental rights against the government's interest in battling terrorism; rather, when there is a conflict, fundamental rights must *always* yield. When the Bush administration talks about how terrorists use our freedoms

against us, the actual policies put in place suggest that the appropriate way to fight terrorism is to emasculate the Constitution's guarantees of basic human rights—though Bush has never stated any such explicit principle. Since the terrorists use our freedoms against us, we must eliminate or severely limit those freedoms—not recast them in ways that make misuse more difficult in the context of the rule of law.

Perhaps we have misunderstood the Bush administration's interpretation of "freedom." In China, in the Middle East, and increasingly in other parts of the world, freedom is being interpreted as "economic liberty," not as political liberty.[31] I do not have irrefutable proof that the Bush administration is going this route, but I do believe it is.[32] The problem is that we will not have such proof until that group is forced to lay its principles out, and show how these principles support its policies—both actual and proposed. As long as the administration can use the "value" of "freedom" to mask its true principles and policies, it will be able to put its own "spin" on what it is doing. People, including the media, will look at the announced "value" and miss the principle(s) for which it stands as proxy. The policies that are attendant are usually so complex that the press is unable or incompetent to show the mismatch between value and policy. Principles are not so easy to manipulate. Their logical relationships with other principles are easier to ascertain. If the Bush administration announced that it stood for the dominance of economic freedom over individual political liberties—in a climate in which many already believe the administration is too friendly with large corporations—it would be very difficult for administration apologists to deflect the criticisms that would come—especially from those who understand our traditional political freedoms, as they are represented in the Constitution.

8. Patriotism, Fear, and "Covenanted Patriotism"

The importance of the administration's emphasis on the value of national security cannot be overemphasized. They tie national security to the value of patriotism; and the administration's underlying principle governing such "patriotism is this: if you are patriotic, you must be for whatever the Bush administration says will keep the nation safe. On the other hand, if you are not walking in lockstep with the demands the administration makes on behalf of national security, you are not patriotic. To charge someone with being unpatriotic is very serious since it is tantamount to a charge of treason.

Abraham Lincoln characterized patriotism in a way that is at odds with the view of the Bush administration. Lincoln argued for what he called "covenanted patriotism."[33] He said *this nation is unique in that it based on an idea embodied in the principles of political freedom found in the Constitution—*

principles that philosophers like Rawls and Dworkin think are not subject to being "weighed" against utilitarian concerns.[34] According to Lincoln, a true patriot thinks of him or herself as bound by a covenant with the other citizens of this nation. That covenant is based on the promise he or she shares with other citizens to defend the nation—*and its defining principles*—from its enemies.[35] That promise, according to Lincoln, guides each citizen's life.

On this view, then, patriotism is based, ultimately, on the *principles* of political liberty embodied in the Constitution—not on the "value" of national security. The citizen patriot is bound to defend the nation, but the important thing to notice is that the concept of "the nation" is circumscribed by the principles of individual political liberty embodied in our founding document. In short, to give up these principles in order to defend the nation is a contradiction. If we give up these principles, we are no longer defending the United States of America. We are defending something, but it is not the country we are covenanted to defend. This idea is captured nicely in the film "Good Night, and Good Luck" about the battle between Edward R. Morrow and Senator Joseph McCarthy. Toward the end of the film, two members of Morrow's team, a married couple, are talking. Joe says to Shirley, "What if we're wrong…An argument could be made for the greater good." Shirley replies, "Not if you give it all away; it's no good then." I take Shirley to be saying that giving away the basic political liberties of this country is too high a price to pay for "national security" because, in that case, you are no longer defending this country.[36]

9. "Family Values" as a "Megavalue"

The problems that arise from operating within a political metaframe of values are especially apparent when single values are replaced by what I will call "megavalues."[37] There are several of these megavalues that have been very important in our political discourse, but the most important in our current political environment is embodied in the term "family values." The full treatment of megavalues must await a follow-up paper; but the importance of this notion requires we at least sketch out how these "clusters of values" are both similar to and different from individual values.

I have argued that the principles associated with a particular value, such as national security, are often very different for people with different political orientations. We saw that this disparity makes usually productive political debate and dialogue difficult or impossible. Megavalues greatly magnify the difficulties associated with single values. To many Americans, "family values" entails things like support for public education, high quality and relatively inexpensive child care, affordable health care insurance for all, access to college and university education for people who are not wealthy, child care services to protect abused children, and so on. The principles associated with

these things are what this first group naturally associates with the term "family values." Those who are privy to the ideology of the present administration understand that, when Bush and those associated with him use the term "family values," they mean things like opposition to abortion, resistance to government support of embryonic stem cell research, defining marriage in a way that precludes gay individuals from being married, and (for many, though not all) making sure that illegal aliens who have been in this country for many years must go back to their country of origin rather than have a chance to achieve citizenship.

Some of the things the Bush administration understands to be part of "family values" are treated by the administration as "trumps"—or, perhaps "secondary trumps." For example, opposition to abortion is clearly the most important "family value" for the current administration. I recently heard a story of a woman who was accused of being against family values. She replied that she had been in a monogamous relationship for over thirty years, had reared two excellent children, was active in her house of worship, and so on. None of this mattered to her accuser, who simply played the abortion card like a bridge player laying down a trump. It was true that the accused woman was "pro-choice," but she based her political stance on a well conceived ethical justification. None of that mattered to the woman accusing her of being against "family values." The abortion issue was trump in her view—and, I suspect, in the view of most people who support the present administration.

Today, being against gay marriage is a sort of "secondary trump." It does not carry the same weight as being opposed to abortion in the minds of people like the one who accused my woman friend of being against "family values"; but it is a close second. The bottom line, then, is that all "family values" are not regarded equally. This allows the administration to be cavalier about many, or even most, of their "family values" as long as they stress the ones that are trumps or secondary trumps. The effect is that the cluster of values that the term "family values" *typically signifies* (in the minds of most Americans) plays a distant second fiddle, in the view of the Bush administration, to the one or two things that the administration regards as the really important ones—in this case, abortion and gay marriage.

It follows that principles and policies relating to many really important issues pertaining to families are never articulated. Instead, the administration uses the mantra of "family values" to suggest it is concerned about families, but ideologues within the Republican Party understand that important policy initiatives will only be associated with abortion, gay marriage, and (perhaps) embryonic stem cell research. The other "family values" will get some lip service, but will never make it to the stage where policy changes are considered. Thus, megavalues like "family values" facilitate a "bait and switch" approach that would not be possible in a metaframe of principles. If a

politician or party is forced to lay out the principles associated with its stress on something like families, it will not be easily possible to waive a semantical wand over the area and then focus on one or two "trump" or "secondary trump" issues; yet that is exactly what a metaframe of values encourages.

10. "Spin" and Political Cynicism

Earlier in this paper (in Section 7), I used the term 'spin', which has become very popular in the media. Both the general public and many of my students believe that "all political discourse is just spin." It follows that there are no political truths, so politics is essentially "just a game." All that matters is who "wins." This leads to political cynicism. People come to believe that all politicians are liars.[38]

While I am uncomfortable claiming that accepting a values metaframe is the primary cause of this phenomenon, embracing such a framework is clearly both consistent with such a view and arguably linked to it. Values are concepts. As I have argued, they do not participate in a web of logical interconnections, as principles do, and we have seen that they are not amenable to being placed in a "values hierarchy" without reference to circumscribing principles. This makes values much more ambiguous than principles. It also contributes to lack of argumentative clash, as we saw above; and this makes it much easier for unscrupulous individuals to utilize the ambiguity inherent in a values metaframe to "suggest" that they mean what their listeners want them to mean. The ideological base of these manipulative individuals and groups understands, but many who might vote against them if they really knew these politicians' intentions will take the path of least resistance and assume these ambiguous values will be fleshed out in terms of the principles they support. When this does not happen, these voters will become political cynics who distrust all politicians.

An example may make this clearer. Academic Stanley Fish has argued that a supposed example of "spin" is really not such an example at all. Presidential advisor Karl Rove had said, according to Fish, that real disposable income has risen 14% since Bush took over the White House.[39] Two critics (Jackson and Jamieson), Fish points out, regard this as spin. Since Rove's statistic was referring to *the whole economy* and since the great majority of the economic gains went only to those at the top of the economic pyramid, Rove's statement was clearly spin. These critics argue that Rove was *implying* that the gains were more evenly distributed among wage earners in the United States. Fish challenges this by pointing out that Rove and the critics simply have different foci: Rove on *growth of the economy as a whole*, even if most of it has gone to those at the top, and his critics on *a more widespread growth in disposable income*. On the basis of my analysis in this paper, I

would argue that Rove is using a *metaframe of values* to mask the truth. The value in this case is "the health of the U.S. economy." His critics, on the other hand, are appealing to a *principle of just distribution*. Rove would probably counter with some sort of "trickle down" *principle*, but most Americans would understand that they had not participated in the economic gains. This would lead many, if not most, to question the viability of Rove's "trickle down" principle. In short, Rove's spin, like a top slowing down, starts wobbling when the *principle* he is appealing to is made clear. Fish may be right that there is no certainty, but philosophers have known that for a very long time. He is clearly wrong, however, in his argument about this example. Most people would (and did) interpret Rove's claim as meaning that Bush had benefited people across the economic spectrum—not just those at the top. On the other hand, the Republican base knew what he was saying. This is completely consistent with my argument in this article.

Forcing politicians to utilize a metaframe of principles means forcing them to make their true intentions clear. There will still be some ambiguity, but people will be much less likely to misinterpret what they are being told. This will hopefully reduce the prevalence of political cynicism and make civic discourse more civil. For example, running "attack" political advertisements regarding principles is much more difficult than running them in a framework of values.

In a metaframe of principles, then, the notion of "spin" has much less impact. Describing an issue in terms of a different set of principles leads to an examination of the implications of these principles and their logical relations with others to which one is committed. Such a scenario leaves little room for the deception and manipulation associated with spin.[40] Someone attempting to recast an issue in the framework of a different set of principles than the one that has been proposed will be forced to confront the logical interrelations of these two sets of principles. In a metaframe of values, on the other hand, as the example with which I began this article shows, it is very easy for political partisans to *recast* an issue by *situating it* in a framework of different values.[41] The standard line for doing this was suggested at the beginning of this paper: "This issue is *about* jobs and economic development for the state of Michigan [not *about* preventing environmental damage or protecting the lower income families that live along the proposed pipeline route]."[42] It will be much easier for "spin doctors" to avoid dealing with another set of values that has been used to frame the issue precisely because the relationships between the two sets of values are difficult or impossible to specify. "Spin" is unmasked as what it really is: deception for the purpose of manipulating the electorate.[43] In a metaframe of principles, such deception is difficult or impossible. Therefore, political cynicism ceases to be a major issue.

11. Conclusion: The Appeal of the Values Metaframe

Though my arguments to this point are strong, they may not be strong enough to carry the day. The metaframe of values has gotten a lot of "traction" in the wake of the 2004 presidential election. People on all sides of the political spectrum have embraced it and will be reluctant to abandon it for a metaframe of principles *unless they can understand why they gravitated to a values metaframe* in the first place.

Referring to "our values" is a very traditional way of representing our ethical positions, but it is fatally flawed. During the 2004 presidential and congressional campaigns, Democrats should not have accepted the "values" political metaframe suggested by Bush strategists. On that turf, the Bush administration was able to mask their principles—and related policies—under a smokescreen of "values." Such a political metaframe of values will typically favor the party in power, especially if they have a measure of media control. The reason for this is that, in a values metaframe, it is usually a matter of which party "gets there first"—that is, which party is able to define their "values" in a favorable way before the other party can. For example, when Republicans were able to make national security "their" value, Democrats were reduced to either saying, "Me too," or to being perceived as being *for* danger. By stressing the importance of preserving civil liberties in the face of a perceived terrorist threat, Democrats appeared to be placing too little emphasis on national security—and being, by implication, too willing to place the nation in peril.

In short, in a values metaframe, the issue of who "wins" and who "loses" becomes a matter of timing. Under such a political metaframe the "challenger" is always at a disadvantage. People tend to think in binary terms; therefore, the challenger will always be viewed as "against" whatever value is under discussion and "for" whatever ill or vice is the opposite of that value. Even if such a challenger can avoid being so portrayed, the contest will devolve into "I'm for national security" and "Me, too" which doesn't appear to create a choice for voters. People will gravitate to the *status quo* under these conditions. If the *status quo* is not available, people will tend to take a conservative approach because it minimizes change. Of course, both approaches at this point in history will favor Republicans.

If, on the other hand, Democrats had defined their positions in terms of *their principles*, in 2004, they might have forced the Bush people to be much more specific about what *they* believed—*and about their true agenda.* The Bush administration would not have been able to mask their true intentions with talk of "red" and "blue" states. That absurdly simplistic dichotomy—which I regard

as a *bogus* political metaframe—would have disappeared once politicians began talking about their principles and how they intended to *balance* them.[44]

Making this change will, I hope, lead to a politics that does not produce a president whose principles and policies are rejected by a majority of Americans, according to polls taken after the 2004 election.[45]

12. Postscript: Unfinished Philosophical Business

This paper includes elements of a rather involved monograph. Because of the current political importance of its leading ideas, however, I have avoided many complications that a more thorough treatment will demand. In this "Postscript" I want to call attention to three of these issues. Some other matters are mentioned in endnotes; and some issues I regard as important to a comprehensive treatment are ignored for the sake of simplicity.[46]

First, there is the matter of what I take principles to be. I said earlier in the paper that values are concepts, like honesty or piety. I alluded to the fact that principles are expressed in sentences and stand in logical relations with one another in ways values cannot. Other than that, I have said little about what principles are. Principles are usually considered to be rules. But the principles I refer to here will be rather deep and general rules—and this is also consistent with much current philosophical usage. In addition, I assume something like Rawls' Wide Reflective Equilibrium (WRE) as the basis for normative ethical justification. WRE encompasses three areas: background assumptions, specific considered moral judgments, and moral principles. All three of these areas must be brought into "equilibrium" in order for justification to be possible. Rawls does not privilege any one of these areas. If there is a conflict between, say, a specific considered moral judgment and a moral principle, one must be abandoned, or modified so that the conflict is no longer manifest; but Rawls does not provide a formula for making such a choice. Instead, he tells us that such determinations must be made on a case-by-case basis.[47]

It follows that there are rules which probably would not qualify as principles simply because they would be too vague and unspecified. In another paper, I distinguish between what I call "Ten Commandment" type principles and those mediated by a Wide Reflective Equilibrium. I argue that the former are usually too vague to be helpful.[48] This distinction is essential to a more complete treatment of these issues. The reason for this is that philosophers are well aware of various "tricks" that can be used to blur the distinction between values and principles I draw here. For example, someone may suggest a template for converting values to principles along the following lines. The value of "honesty" could be rendered as the rule: "Be honest" or "One should always be honest."

This, of course does little more than state that one should adopt the value of honesty. If, on the other hand, something like the WRE model were adopted, such simplistic conversions would run afoul of the requirements of that model. For example, particular moral judgments would have to be brought to bear to clarify the "principle," as would various background assumptions, which would make the context in which such a principle is deployed more apparent. A "principle" as vague as "Be honest" would be virtually impossible to bring into equilibrium with other principles, specific considered moral judgments, and background assumptions.[49]

The second issue that a more complete treatment of this matter would address is the relation between justification and explanation with regard both to values and to principles. Economists are widely regarded as the most "scientific" of the social scientists, primarily because of their use of very detailed mathematical models. Some political scientists and sociologists also utilize such models. Economists' detailed models regularly invoke such "principles" as that of "maximizing expected value," where the term "value" is taken to be a measure of wanting. As a philosopher of science, I have problems with some of the things economists, and other mathematical social scientists, do. I have written about some of these.[50] Nevertheless, the fact that principles are logically related, the proclivity of principles to more precise formulations, and their ability to justify decisions make them much better candidates for theory building in economics and the other social sciences than vague references to values.[51]

Philosophers of social science called "interpretavists" also invoke principles in their explanations, but for a very different reason. These theorists note that principles often make actions possible by locating them in a defining framework.[52] They note, for example, that it would not be possible to "slide into third base" without the game of baseball and its comprehensive set of rules and principles. Similarly, Rawls notes that visiting judicial punishment on someone for committing a crime is impossible without the criminal justice system and its circumscribing legal principles. These observations suggest that principles may well be essential to the very performance of certain actions. If that is so, such principles will certainly play a role in explaining these actions.

I propose that:

> Normative decisions are often best justified by using principles that are also capable, in theory, of contributing to the explanations of such decisions.

I call this proposed standard "The Justification/Explanation Rule." The attempt to use "values" as the basis for such judgments does not meet the test implicit in the justification/explanation rule.[53] Because of their inherent vagueness, values can be used to explain a

variety of different, and often conflicting, decisions. Their greater precision and logical interconnections make principles better candidates for theory building—especially when our concern is with actions that have a normative dimension. Their accuracy also makes them more likely to satisfy the justification/explanation rule. This is, however, a very complex issue; I suggest it here in part because it relates directly to my final point.

This third issue is "virtue ethics." The most widely held versions of virtue ethics appear to assume that people have some very broad-based dispositions and that these can be used both to justify their decisions and to explain them—thus satisfying the justification/explanation condition I alluded to earlier. Recently, Gilbert Harmon has argued that standard versions of virtue ethics (not including Hume's version) run afoul of recent work in social psychology.[54] Harmon claims the notion that people have these broad-based dispositions is found in the area of psychology called personality theory, which Harmon claims is a discipline in disarray. The late Robert Solomon, in a companion piece, challenges Harmon's claims. I cannot resolve this complex matter here, but we should be aware of it. Obviously, if there really are broad-based dispositions and if these enter into explanations of people's actions, then there may be somewhat more to a "values metaframe" than this paper suggests—though I believe most of my conclusions would survive. In an extended version of this work, I will defend Harmon's position regarding "virtue ethics." This is critical since virtue ethics is currently very important in areas of "applied ethics" such as business ethics. Considerations that suggest a metaframe of principles is superior to one of values also mitigate strongly against a "virtue ethics"—either for use in justifying normative claims or for theory building in psychology or the social sciences.

Notes

[1] George Lakoff, *don't think of an elephant* (White River Junction, VT: Chelsea Green Publishing, 2004). See especially the preface and the first chapter. Lakoff's concept of a frame is very rich, but his position regarding values is consistent with what I speak of here, as the first chapter makes clear. Note that I have also used the concept of a frame independently. See my article, "Winning in the Court of Public Opinion," *The Romeo Observer* 12 June 2002: 6-A.

[2] Thomas P. Morgan, "Is it Environmental Discrimination," *City Pulse* 8 June 2005: 3. My italics.

[3] Lakoff, especially the preface and first chapter.

[4] Lakoff, *seriatim*, especially pp. 117-118.

[5] In section 12, "Postscript: Unfinished Philosophical Business," I take up the issue of how principles can be filtered through what Rawls calls a "wide reflective equilibrium." Undoubtedly, such an analysis adds important nuances to this study. It is not, however, essential to the central ideas of the paper.

[6] *The New York Public Library Desk Reference, Third Edition* (New York: The Macmillan Co., 1998) 849.

[7] Jess Brown, "Justices Bar Guantanamo Tribunals." *The Wall Street Journal* 30 June 2006: A9.

[8] Some might object to my strong focus on the Bush administration in this paper—suggesting that similar things could be said about any previous administration. There are two reasons for my focus on the current administration. First, the Bush administration has refined the "art" of "values framing" far beyond anything seen in the past and used such framing to threaten the form of government embodied in the U.S. Constitution. Second, this is happening now and it is vitally important that individuals (including scholars) understand how these threats to the Constitution are being framed. Such understanding is essential if these threats are to be confronted both now and in the future. This is not a history paper. It points out what I regard as a very serious problem and suggests how to deal with it.

[9] A. H. Maslow, *Motivation and Personality*, 2nd Edition (New York: Harper & Row, 1970). Maslow's work focuses on a hierarchy of *needs*, but this suggests that the corresponding values might also form a hierarchy.

[10] See, for example, Amartya Sen, *Rationality and Freedom* (Boston: The Harvard UP, 2003) for material on multidimensional comparisons.

[11] Margaret Talbot, "The Way We Live Now" *New York Times* 28 September 2003, mag. desk.

[12] Talbot.

[13] Eric Lichtblau and Scott Shane, "Bush Is Pressed over New Report on Surveillance," *New York Times* 12 May 2006, national desk, late edition – final: A1.

[14] In 2005, the House of Representatives rejected *some* of the most "Constitutionally challenged" parts of the U.S.A. Patriot Act, which had to be reexamined by Congress because of sunset provisions in the original act. Bush threatened to veto any version of the Act which did not contain the strongest invasions of U.S. civil liberties—for example, those allowing examining library records and bookstore purchase data without a judge's approval. See Carl Hulce, "House Blocks Provision for Patriot Act Inquiries," *New York Times* 16 June 2005: National Desk: http://select.nytimes.com/search/restricted/article?res=F60613FE3C5F0C758DDDAF0894DD404482.

[15] R. W. Apple Jr., "After the Attacks News Analysis; No Middle Ground," *New York Times* 14 September 2001, late ed. – final: A1.

[16] The "fallacy of faulty dilemma" is essentially a fallacy of false premise. In the context of a dilemma (viewed as a type of argument), there is a premise that states that either A or B is the case. The fallacy occurs because there is some viable third alternative not identical with either A or B. For example, if someone says a cup of coffee is either very hot or very cold, this fallacy is arguably present. The coffee may, in fact, be warm but not very hot. Similarly, someone may support many aspects of the U.S. foreign policy but fail to support going to war in Iraq. Such a person is not properly characterized as being "against us" in any appropriate sense.

[17] Sharon Begley, "Our Brains Strive to See Only the Good, Leading Some to God," *The Wall Street Journal Online* 28 October 2005: The Science Journal.

[18] See especially the first chapter of Paul Krugman, *The Great Unraveling* (New York: WW Norton and Co., 2002).

[19] The sort of balancing I am suggesting might reveal different ways to achieve a given level of national security, and these ways might be ascertained by examining the relevant principles. Different principles might produce the same level of national security but have negative/positive impacts on other values. As my friend David Zin, an economist, points out, this is analogous to applying the positive/negative economic externality concepts to principles. I am grateful to Mr. Zin for many helpful comments about this work.

[20] Ronald Dworkin, *Taking Rights Seriously* (Cambridge, MA and London, England: The Harvard UP, 1977). Also see John Rawls, *A Theory of Justice* (Cambridge, MA: Belknap Pr, 1971) and Justine Burley, ed., *Dworkin and His Critics* (Malden, MA and Oxford, UK: Blackwell Publishing, 2004).

[21] George W. Bush, "Bush's Veto Message," *New York Times* 19 July 2006: http://www.nytimes.com/2006/07/19/washington/text-stem.html. In this speech, Bush tacitly affirms the original principle stated in the text, but he also states a more general principle: that we should embrace technology without becoming slaves to it. What is interesting about this new principle is that it is much more vague and ambiguous than his first principle (that we should not take life to preserve life). In effect, this new "principle" says technology is a value but it has limits. In short, Bush has morphed his initial principle into something that has the characteristics we have argued make values dangerous.

[22] This is true even if we accept Ted Koppel's analysis that the U.S. now needs to stay in Iraq to block Iran. This follows from the fact that Saddam Hussein was doing that. Bad as he was, he was blocking Iran.

[23] There are many examples of the Bush administration adopting policies that seem based on the principle that the end justifies the means, but the most prominent is clearly Iraq. Absent weapons of mass destruction, the justification for the carnage in Iraq must be based explicitly on the idea that the end justifies the means. Many, of course, do not believe that the end, whatever it is, does justify the mess that Iraq has become.

[24] James E. Roper and Timothy W. Sommers, "Debate as a Tool for Teaching Business Ethics," *In the Socratic Tradition: Essays on Teaching Philosophy,* Tziporah Kasachkoff, ed. (Lanham, MD: Rowman and Littlefield, 1998): 91-103.

[25] Roper and Sommers, 91-103.

[26] *Why We Fight*, Dir. Eugene Jarecki, Sony Pictures Classics, 2006.

[27] See Nolan McCarty, et al. *Polarized America* (Cambridge, MA: The MIT Pr, 2006).

[28] Neil A. Lewis and Robert Pear, *New York Times* 5 October 2001, national desk. The following quotation from this article is adequate warrant for the claim I make in the text: "The Senate is often thought to be firmer than the House in preserving civil liberties protections. But the House Judiciary Committee's greater efforts on this issue appear to have been a result of complaints about the administration's approach not only from liberal Democrats but also conservative Republicans."

[29] See James E. Roper, "Government is Hypocrisy on Stilts," *The State News,* December 7, 2006: 4A. In this article, I point out that Ron Suskind, in *The One Percent Doctrine*, tells us that the "Cheney Doctrine" specifies that "if there was even a one percent chance of terrorists getting a weapon of mass destruction…the United States must now act as if it were a certainty. This was a mandate of extraordinary breadth." I go on to make clear that Cheney is not necessarily talking about a situation in which everyone is killed; "WMD" can mean many things—including the anthrax attack on the

U.S. mail. This article is a useful addition to this paper because, in it, I contrast the administration's approach to terrorism via this "Cheney Doctrine" with its approach to global warming, which carries incredible risks. The Bush administration, nevertheless, has been willing to ignore those risks on the grounds that global warming has not been completely proven.

[30] Carl Hulce, "House Blocks Provision for Patriot Act Inquiries," *New York Times* 16 June 2005: National Desk: http://select.nytimes.com/search/restricted/article?res=F60613FE3C5F0C758DDDAF0894DD404482 .

[31] Manual G. Velasquez, *Business Ethics Concepts and Cases,* 6th ed. (Upper Saddle River, N.J.: Pearson Prentice Hall, 2006). Velasquez discusses the different meanings that are often attached to the terms "justice" and "freedom" in the context of business—meanings that associate these concepts with economic liberty.

[32] Robert Wright, "An American Foreign Policy That Both Realists and Idealists Should Fall in Love With," *New York Times* 16 June 2006. This article says, "Free markets are spreading across the world on the strength of their productivity, and economic liberty tends to foster political liberty."

[33] John H. Schaar, *Legitimacy in the Modern State* (New Brunswick (U.S.A.) and London (U.K.): Transaction Books, 1981): 285-312.

[34] See note 20 above.

[35] James E. Roper, "Market Failure, Symbolic Meaning, and the Covenant of Democracy," *International Journal of Ethics* 3.3 (October 2004): 321-335.

[36] *Good Night, and Good Luck*, Dir. George Clooney, Warner Brothers, 2005. It is important to note that Bush has repeatedly said his "War on Terror" will probably last for a very long time. That means that we will not be "giving it all away" for a short period. We will be "giving it all away" indefinitely, and that suggests we will not be getting it all back—ever.

[37] The term "megavalues" was suggested by David Zin referred to in earlier notes.

[38] Stephen Nathanson, *Should We Consent to be Governed?* (Toronto, Canada: Wadsworth, 2001) 27-45.

[39] See Stanley Fish, "The All-Spin Zone," *New York Times* 13 May 2007: Opinion (on-line). The remainder of this paragraph is based on Fish's article.

[40] Note that context can also be an issue when working with principles, but it is typically much more important when working in a metaframe of values. Rawls' "equal liberty" principle might be interpreted as applying only to economic rights. That was not Rawls intent, and this can be ascertained by a careful reading of Rawls' text; but it may be tempting to some to interpret Rawls in this way—placing this principle in a context much different from the one intended by Rawls. As I

said, this can be checked by studying Rawls' text. It is often much more difficult to determine that a value has been placed in a context that dramatically alters its meaning because values are much more vague than principles—and, of course, they lack the logical interconnectedness of principles.

[41] I recently saw a bumper sticker that said "Choose Life." If the "values" are "choice" and "life," then "choose life" represents a way of "spinning" the issue toward the value of life. In a sense, this bumper sticker "captures" the idea of choice by recasting it into a context where life is the dominant value. "You want 'choice', O.K. then choose life—not death." In a political metaframe of principles, though, this "bumper sticker politics" just seems incomplete. What is the context of the "choice"? Whose "life" are we invited to choose—the mother's, the fetus', some third party's? The most important considerations in such a framework are the *principles* associated with the various "abortion" positions—and the logical connections among these and other accepted principles. For example, what does the term 'life' mean in the context of the various principles this web of issues evokes?

[42] See note 2 above. Additional italics are mine.

[43] The philosopher Harry Frankfurt has written an illuminating little book called *On Bullshit* (Princeton and Oxford: Princeton UP, 2005). The key point Frankfurt makes about his subject is that bullshit is not truth seeking. Even a liar has to acknowledge that there is a truth. The point of a lie is to try to get someone to believe to be true what the individual telling the lie knows to be false. Bullshit has no such allegiance to the truth. Its goal is to manipulate the listener by using words. The speaker may intimate that what he or she is saying is true, but the speaker really doesn't care whether it is or not—only that it has the intended effect on the listener. A case can be made that spin is bullshit. A metaframe of values turns out to be a superb tool for someone who is "bullshiting."

[44] In the Special Features section of the film *Why We Fight*, director Eugene Jarecki answers questions in a section called Q & A. The first question he answers is "What do you hope to achieve [with this film]?" His answer refers to the work of Princeton University cartographers who have shown that the United States is not really composed of "red" and "blue" states, but is really a mixture, which he shows on the screen. Jarecki argues (correctly, I believe) that this simplistic dichotomy is very damaging to the political discourse of the United States.

[45] Robin Toner and Majorie Connelly, "Bush's Support on Major Issues Tumbles in Poll," *New York Times* 17 June 2005: National Desk: http://select.nytimes.com/search/restricted/article?res=F30E12FC3F5F0C748DDDAF0894DD404482.

[46] My standards for a "thorough" treatment are very high, and include many deep philosophical issues that would not normally be discussed in an introductory essay.

[47] Nelson Goodman first introduced this idea in connection with justifying principles of inductive reasoning in *Fact, Fiction, and Forecast, Fourth Edition* (Cambridge, MA and London, England: Harvard UP, 1983) 60, 63-64.

[48] See James E. Roper, "A Philosophical Perspective on Corporate Codes of Ethics," *Research in Ethical Issues in Organizations, Volume 6,* ed. Moses L. Pava and Patrick Primeau (London: Reed Elsevier, 2005) 195-206.

[49] My economist friend David Zin, referred to earlier, makes a very interesting point about WRE in relation to this paper. Zin points out that the Articles of Confederation and the Constitution are both based on similar *values*; but these documents have very different meanings in light of the very different *principles* embodied in the two documents. People do like to summarize their views in terms of their values. Zin's point is that such summaries are hopelessly vague and ambiguous without a set of clarifying principles.

[50] James E. Roper and David Zin, "Review of Amartya Sen, *Rationality and Freedom,*" *Essays in Philosophy*, June 2004. http://www.humbolt.edu/-essays/roperrev.html.

[51] Note that the principle of maximizing expected value refers to "value." In fact, the "values" in question are really preferences among alternatives some of which are basically lotteries among different possible payoffs. So such references to "value(s)," in the context of the principle of maximizing expected value, are really probabilistically weighted versions of someone's preferences among different alternatives. The principle of maximizing expected value utilizes these weighted preferences to determine a "rational" choice in a specified decision situation.

[52] Merrilee H. Salmon, "Philosophy of the Social Sciences," *Introduction to the Philosophy of Science,* edited by Merrilee Salmon *et al.* (Englewood Cliffs, N.J.: Prentice Hall, 1992). Ch. 11.

[53] James E. Roper, "How is Rational Decision Theory Possible?" Delivered to a Philosophy Colloquium, at Michigan State University, 1974.

[54] Lisa Newton, ed., *Taking Sides: Clashing Views in Business Ethics and Society, 9th Edition,* Dushkin/McGraw-Hill, 2005. Issue 2 (Virtue Ethics).

Works Cited

Apple Jr., R. W. "After the Attacks News Analysis; No Middle Ground." *New York Times* 14 September 2001, late ed. – final: A1.

Begley, Sharon. "Our Brains Strive to See Only the Good, Leading Some to God." *The Wall Street Journal on Line* 28 October 2005: The Science Journal http://online.wsj.com/article/SB113045290939681793.html.

Burley, Justine, Ed. *Dworkin and His Critics*. Malden, MA and Oxford, UK: Blackwell Publishing, 2004.

Brown, Jess. "Justices Bar Guantanamo Tribunals." *The Wall Street Journal* 30 June 2006: A1, A9.

Bush, George W. "Bush's Veto Message." *New York Times*. 19 July 2006: http://www.nytimes.com/2006/07/19/washington/text-stem.html.

Dworkin, Ronald. *Taking Rights Seriously*. Cambridge, MA and London, England: The Harvard UP, 1977.

Fish, Stanley. "The All-Spin Zone." *New York Times* 13 May 2007: Opinion. http://fish.blogs.nytimes.com

Frank, Harry G. *On Bullshit*. Princeton and Oxford: Princeton UP, 2005.

Good Night, and Good Luck. Dir. George Clooney. Warner Brothers, 2005.

Goodman, Nelson. *Fact, Fiction, and Forecast*, 4th Edition. Cambridge, MA and London, England: The Harvard UP, 1983.

Hulce, Carl. "House Blocks Provision for Patriot Act Inquiries." *New York Times* 16 June 2005: National Desk.

http://select.nytimes.com/search/restricted/article?res=F60613FE3C5F0C758DDDAF0894DD404482

Krugman, Paul. *The Great Unraveling.* New York: Norton WW and CO, 2002.

Lakoff, George. *don't think of an elephant.* White River Junction, VT: Chelsea Green Publishing, 2004.

Lewis, Neil A. and Robert Pear. *New York Times*. National Desk. 5 October 20. http://select.nytimes.com/search/restricted/article?res=F4081EFC3C580C768CDDA90994D9404482.

Lichtblau, Eric and Scott Shane. "Bush Is Pressed Over New Report On Surveillance." *New York Times* 12 May 2006 national desk, late edition – final: A1.

Maslow, A. H. *Motivation and Personality*, 2nd Edition. New York: Harper & Row, 1970.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, MA: The MIT Pr: 2006.

Morgan, Thomas P. "Is It Environmental Discrimination?" *City Pulse* 8 June 2005: 3.

Nathanson, Stephen. *Should We Consent to be Governed: A Short Introduction to Political Philosophy, 2nd Edition.* Toronto, Canada: Wadsworth, 2001.

The New York Public Library Desk Reference, Third Edition. New York: The Macmillan Co., 1998.

Newton, Lisa, ed. *Taking Sides: Clashing Views in Business Ethics and Society, 9th Edition.* Dushkin/McGraw-Hill, 2005.

Palast, Greg. *Armed Madhouse: Who's Afraid of Osama Wolf? China Floats, Bush Sinks, The Scheme to Steal '08, No Child's Behind Left*. New York: The Penguin Group, 2006.

Rawls, John. *A Theory of Justice.* Cambridge, MA: Belnap Pr, 1971.

Roper, James E. "Government is Hypocrisy on Stilts." *The State News,* Thursday, December 7, 2006: 4A.

Roper, James E. "A Philosophical Perspective on Corporate Codes of Ethics." *Research in Ethical Issues in Organizations, Volume 6.* Moses L. Pava and Patrick Primeau, eds. London: Reed Elsevier, 2005: 195-206 (Chapter 11).

Roper, James E. and David Zin. "Review of Amartya Sen, *Rationality and Freedom.*" *Essays in Philosophy*, June 2004. http://www.humbolt.edu/-essays/roperrev.html

Roper, James E. "Market Failure, Symbolic Meaning, and the Covenant of Democracy." *International Journal of Ethics.* Vol. 3, No. 3 (October 2004): 321-335.

Roper, James E. "Winning in the Court of Public Opinion," *The Romeo Observer* 12 June 2002: 6-A.

Roper, James E. "When a 'W' is not a 'W'." *Florida Philosophical Review.* Volume I, Issue 2 (200l): 57-71. http://www.cas.ucf.edu/philosophy/fpr

Roper, James E. and Timothy W. Sommers. "Debate as a Tool for Teaching Business Ethics." *In the Socratic Tradition: Essays on Teaching Philosophy.* Tziporah Kasachkoff, Ed. Lanham, MD: Roman and Littlefield, 1998.

Roper, James E. "How is Rational Decision Theory Possible?" Philosophy Colloquium, Michigan State University, 1974.

Salmon, Merrilee H. "Philosophy of the Social Sciences." *Introduction to the Philosophy of Science. E*dited by Merrilee Salmon *et al.* Englewood Cliffs, N.J.: Prentice Hall, 1992. Ch. 11.

Schaar, John H. Legitimacy in the Modern State. New Brunswick (U.S.A.) and London (U.K.): Transaction Books, 1981.

Sen, Amartya. *Rationality and Freedom.* Boston: The Harvard University Press, 2003.

Suskind, Ron. The *One Percent Doctrine: Deep Inside America's Pursuit of its Enemies Since 9/11.* New York: Simon & Schuster, 2006.

Talbot, Margaret. "The Way We Live Now." *New York Times* 28 September 2003, mag. desk: http://select.nytimes.com/search/restricted/article?res=FB0711FC3E5A0C738DDDAB0894DE404482.

Toner, Robin and Marjorie Connelly. "Bush's Support on Major Issues Tumbles in Poll." *New York Times* 17 June 2005: National Desk: http://select.nytimes.com/search/restricted/article?res=F30E12FC3F5F0C748DDDAF0894DD404482.

Velasquez, Manual G. *Business Ethics Concepts and Cases,* 6th Edition. Upper Saddle River, N.J.: Pearson Prentice Hall, 2006.

Why We Fight. Dir. Eugene Jarecki. Sony Pictures Classics, 2006.

Wright, Robert. "An American Foreign Policy That Both Realists and Idealists Should Fall in Love With." *New York Times* 16 June 2006. http://www.nytimes.com/2006/07/16wright.html?th&pagewanted=print

Omniscience and the Identification Problem

Robert Bass, *Coastal Carolina University*

I once came across a Mark Twain story in which a character said something to the effect that the one thing God didn't know was that he was not all-knowing. As an argument against omniscience, Twain's one-liner doesn't amount to much. Thinking about it, however, led to the kind of puzzles I explore here. Some puzzles about omniscience are connected to other issues, such as whether all claims about the future presently have truth-values. Those in turn are connected to deep issues in the metaphysics of time. (Is the future real, and, if so, in what sense?) Others are connected to questions about knowledge by acquaintance[1]—such as whether God must, in order to be omniscient, know what it is like, say, to be guilty or to have a limited perspective, and whether God can know such things without actually *being* guilty or *having* a limited perspective.

My concern is with a different kind of puzzle, having to do with propositional knowledge, knowledge of facts that can be represented by that-clauses in sentences such as 'John knows that the world is round.' I shall focus upon questions about a supposedly omniscient being who propositionally knows the truth about all current states of affairs.[2] I shall argue that there is no such being.

Since we are speaking about a being who supposedly knows all facts, it is useful to say something about knowledge. We can do this sufficiently for present purposes, despite the absence of a consensus among philosophers on a general account of knowledge. For there are two points upon which there is a consensus or near-consensus—that knowledge is more than true belief and that the 'more,' whatever its precise character, makes the belief *non-accidentally* right. The difference between the knower and the one who truly believes the same fact is not just a matter of luck. It is *no accident* that the knower got it right; it is an accident that the one who merely believes the truth got it right.

Omniscience Alone

With that background, consider omniscience. We can abstract from questions about the relation of omniscience to other divine attributes by considering what omniscience would mean for an otherwise unremarkable being. We can return later to questions about God's omniscience. For now, let us speak of the supposedly omniscient *Todd*.

We can begin with a simple statement of the supposition about Todd:

(1) Todd is omniscient.

If Todd is omniscient, then, for any proposition, p, the following conditional must be true:

(2) If p, then Todd knows that p.

That is, for any truth, Todd knows it. Now consider *hiders*, whose defining property is that they are capable of perfectly concealing their existence from any other being. When a hider conceals its existence, it is in hiding. Plainly, if there are any hiders in hiding, Todd is not omniscient, for then there is a fact that Todd does not know, namely, that there is at least one hider in hiding.[3] He does not know that there is any hider in hiding because he cannot distinguish the state of affairs in which there are no hiders from the state of affairs in which there are hiders in hiding.

Further, if hiders are even *possible*, Todd is not omniscient, for again he would be unable to distinguish between the non-existence of hiders and the existence of hiders in hiding. The possibility of hiders implies that there may be hiders in hiding. Since hiders in hiding are perfectly concealed from Todd, there will be a fact that Todd does not know, namely, whether there are any actual hiders or not.

Still further, even if hiders are impossible, if Todd does not know they are impossible, then he is not omniscient. First, there will then be a fact that he does not know, namely, that hiders are impossible, and second, given that he does not know they are impossible, they will be possible as far as he can tell. He will not be able to distinguish their unknown impossibility from their possibility, and so, will not know whether there are hiders or not.

Thus, if Todd is omniscient, he must know that hiders are impossible, for only if hiders are impossible, and known by Todd to be impossible, can it be the case that Todd is, and knows that he is, omniscient. If hiders exist, or are possible, or are not known to be impossible, then some of Todd's beliefs will be, at best, accidentally true, and therefore not cases of knowledge. Thus, we have

(3) If Todd does not know that hiders are impossible, then Todd is not omniscient,

and its contraposition,

(4) If Todd is omniscient, then Todd knows that hiders are impossible.

Finally, given (1), that Todd is omniscient, the following can be derived:

(5) Todd knows that hiders are impossible.

Of course, that argument can be run in reverse. If hiders are either possible, or else impossible in some way that Todd doesn't know, then Todd doesn't know hiders are impossible. If he doesn't know that hiders are impossible, then Todd is not omniscient. The question that faces us here is how Todd knows—if he does—that hiders are impossible.

Since, if Todd is omniscient, he must know that hiders are impossible, there are two matters to be clarified, having to do with how hiders might be impossible and with how some being might know of that impossibility.[4] We can begin with the impossibility of hiders. They might be impossible in either of two ways. They might be intrinsically impossible, in the way that round squares are. Alternatively, hiders might be impossible, given some other fact or facts—that is, *extrinsically* impossible. The existence of hiders might be excluded by some other fact, in the way that irresistible forces are impossible in worlds containing immovable objects. On the first alternative, hiders could not adequately be characterized in a non-contradictory way, and thus there would be no hiders in any logically possible world. On the second alternative, though hiders would be logically possible, and therefore present in some possible worlds, there would be some proper subset of possible worlds in which some fact or set of facts[5] excludes the existence of hiders. There will be no hiders in any of the worlds within the subset.

If hiders are intrinsically impossible, that will only be of use to Todd if he knows of their intrinsic impossibility. If hiders are, in some possible worlds, extrinsically impossible, that will be of use to Todd only if he knows that he is in one of those worlds. So, *are* hiders either intrinsically or extrinsically impossible? And, if so, how does Todd know?

Before proceeding further, consider Todd's knowledge. If Todd knows that hiders are impossible, that knowledge will be either inferential or non-inferential. Could Todd non-inferentially know that hiders are impossible? There are two reasons for suspicion. First, so far as Todd's non-inferential knowledge is being modeled upon ours, plausible cases of non-inferential knowledge typically yield less than certainty. I do not infer from premises that the person with the familiar profile and gait that I see walking across campus is my friend, David. But, upon catching up with him, I can be surprised to discover that it is someone else entirely. Further, even where non-inferential processes appear to yield certainty, as in the case of whether two straight lines can enclose

a space, the certainty may turn out to be only *certitude*, the state of *feeling certain* that something is so, which may not in fact be so. If the normal products of non-inferential belief-forming mechanisms deserve to be called knowledge, that is due to the fact that those mechanisms are generally reliable. General reliability is not enough for Todd to rule out the existence of hiders, however. For if he reaches the belief that hiders are impossible only through a generally reliable mechanism, and if he also knows, as he must if he is omniscient, that the way in which he comes to the belief is only generally reliable, it will remain possible that, though his belief that hiders are impossible was the product of a generally reliable belief-formation mechanism, it is still mistaken.[6]

Perhaps, however, it is not fair to model the non-inferential knowledge of a being who is supposedly omniscient upon ours. There is still reason for suspicion about claims that Todd has non-inferential knowledge of the impossibility of hiders, if nothing else is provided. This derives from the nature of the case. We are speaking of whether hiders are intrinsically or extrinsically impossible. If hiders are intrinsically impossible, there are no hiders in any logically possible world. If hiders are extrinsically impossible in some world, then that world has some hider-excluding feature. That is, it is not logically possible for there to be a hider in that world, given some other fact that obtains there. If hiders are logically impossible in either way, that will be the kind of thing that can be set out in an explicit argument, and, if Todd is omniscient, that argument will be available to him. Whether or not he is somehow capable of non-inferentially reaching the conclusion that hiders are impossible, there will be available, in principle, an explicit proof of their impossibility, and so, the non-inferential access to the truth that hiders are impossible will not be necessary for Todd's knowledge.

On the face of it, no such proof is available. There is no overt inconsistency, nor does there appear to be any covert inconsistency buried in the notion of a being capable of concealing itself from all others. Though it might be said that the inconsistency is buried too deeply for us to unearth it, that is surely a desperate move, for it amounts to saying that, so far as we can tell, the notion *is* consistent. (And further, the idea of an inconsistency buried too deeply to be unearthed bears as much against omniscience as against being a hider.)

Matters seem no different if we focus upon extrinsic impossibility. It is not that there is some difficulty in conceiving of some fact which is both distinct from and incompatible with the existence of hiders. For example, there might be *finders*, from which no being can successfully hide. If there are any finders, there can be no hiders. More to the point, if there are any omniscient beings, there can be no hiders. The problem with these is that if Todd is to know that there are any finders or any omniscient beings, he must *already* have ruled out the existence of hiders. Appealing to the existence of finders or omniscient beings in the attempt to show that there are no hiders

would beg the question, at least, unless there is some independent proof of their existence. But no independent proof could be satisfactory that did not rule out the existence of hiders.[7] That leaves us with an exacerbated version of the earlier problem: It appears that the existence of something incompatible with the existence of hiders can only be known if it is already known that there are no hiders—which leaves us, in order to show that hiders are extrinsically impossible, trying to show that they are impossible in some other way—that is, presumably, that they are intrinsically impossible. But if that could have been shown, we would never have needed to explore the question of their extrinsic impossibility at all.

God's Omniscience and the Identification Problem

It appears, then, that Todd does not know that he is omniscient and therefore is not omniscient. Do matters change if we consider *God* rather than Todd, that is, if we suppose a being with the characteristics, other than omniscience, traditionally ascribed to God? Will a being who, in addition to vast knowledge, is omnipotent, self-existent, and so on be able to rule out the possibility of hiders?

Much remains the same. There is still the fact that the conception of a hider appears consistent, and if it is, hiders are not intrinsically impossible. Perhaps, however, there is something God knows that would enable him to show that hiders are extrinsically impossible. One suggestion derives from consideration of God's role as creator. The argument would be that there could be hiders only if God had created them, but since he did not, and knows that he did not, there are none. The problem with this is that, according to orthodox conceptions of God, there is at least one thing that exists without having been created by God, namely, God himself. God cannot argue that it is impossible for there to be anything uncreated by God because that would rule out his own existence. And that means that hiders might be beings uncreated by God and, being in hiding, unknown to God.

Perhaps a revision of this line of thought will succeed. God is self-existent—that is, does not depend for his existence upon anything else. If God had a proof that there could be only one self-existent being, he could argue that if hiders exist without being his creations, then they would have to either be self-existent or the creations of some other self-existent being. But since there are no other self-existent beings—which, *ex hypothesi*, has been proven—there is no way for hiders to exist.

Let's set this attempt aside, for the moment, to consider another. Suppose that some (modal) version of the ontological argument is sound. Then, it will be necessarily true that God

exists with the full set of theistic attributes. That being than which no greater can be conceived will be omniscient, omnipotent, perfectly good, self-existent, and so on. But if the ontological argument is sound, God will understand that it is, and therefore will have a proof that there is an omniscient being and hence that there are no hiders.[8]

Both of the foregoing attempts depend upon proofs which have not actually been given, and, in the absence of the proofs, it is difficult to be confident of the solution. In addition, however, they share another feature—both are susceptible to *the identification problem*. In each case, we are supposing that there is a proof of the existence of some being with one or more of the traditional attributes of God. Then, the conclusion of that proof is used as a premise for a further argument to rule out the existence of hiders. But what about the being considering the proof? What is to identify *that* being—call it the *Arguer*—as the one proved to exist?

To spell out the problem, let us call the being supposedly proved to exist a *God-like being*. Then, the general form of both arguments—and, so far as I can see, of any others to similar effect—is this:

(1) There is a God-like being.
(2) If there is a God-like being, it knows that it is omniscient.
(3) If some being knows that it is omniscient, then it is omniscient.
(4) The Arguer is the God-like being.
(5) Therefore, the Arguer is omniscient.

That is valid, so let us consider whether the premises can be known to be true. For the first premise, we are supposing that there is an independent proof. The second may require the satisfaction of additional conditions, for example, that the God-like being has not created any hiders, but let us suppose that whatever conditions are needed for it to be true are also satisfied. The third premise is analytic, given that a minimum condition for knowledge is truth. The fourth premise, however, raises the identification problem. How would the Arguer know that he is the God-like being? If the Arguer does not know that, he does not know the conclusion (and, of course, the conclusion would then not be true). More concretely, the problem can be illustrated from the two arguments considered above that depend upon demonstrations of the existence of a God-like being.

Consider first the proof of the uniquely self-existent being. How will the Arguer know that he is himself the uniquely self-existent being? Perhaps, he could have a memory of an infinite past, but that is not enough. Even if there is a uniquely self-existent being, the Arguer might have always been dependent upon and, unknown to itself, sustained in existence by some other being. And if so,

the real uniquely self-existent being might be both a hider and the one upon whom the Arguer depends.

Or consider the ontological argument. The conclusion of the ontological argument, if it is sound, guarantees that there is an omniscient being and therefore that there are no hiders. But what is to certify to the Arguer that he is the omniscient being proved to exist? One suggestion is that the Arguer might know himself to possess some other property, also proven by the ontological argument to uniquely characterize God. But that leads to a dead end for two reasons. First, whatever the ontological argument can prove, it is not clear that it proves that a being than whom no greater can be conceived has no equals in *any* respect.[9] There seems no logical impossibility in omniscient God creating omniscient Todd, provided, of course, that omniscience itself is logically possible. And if the divine attributes can be exemplified in part by other beings, it will not be possible for the Arguer to mount a sound argument that, because he has divine attribute *A*, he must also possess the other divine attributes, *B*, *C*, and so on.

Second, there is a generalized version of the identification problem. In addition to the problem involved in a being knowing herself to be omniscient, it is not clear how she could know herself to have any of the divine attributes. How, for example, would an all-powerful being know that she was all-powerful? It might, of course, be that she had never found any difficulty in executing her will, but that does not show that she is all-powerful, as distinct from never having tried anything that exceeded her powers. For that matter, the experience of never being frustrated in executing her will does not even show that she has not *already* attempted something beyond her power, for she may have been the beneficiary of concealed assistance. There is no apparent way for an all-powerful being to know that she is all-powerful, for an omnipresent being to know that she is omnipresent, for a perfectly good being to know that she is perfectly good, for an indestructible being to know that she is indestructible, and so on. What that means is that the attempt to use knowledge of the set of traits allegedly proven to exist in combination by the ontological argument as a solution to the identification problem is not going to work: The identification problem will have to be solved for the other traits, too.

Thus, the omniscience problem, rather than being solved by appeal to the ontological argument, actually provides evidence that the ontological argument is *not* sound. For, if it were sound, then there would be an omniscient being, and therefore hiders would be, and could be known to be, impossible. But it appears that any being considering this argument and faced with the identification problem would have to recognize that he might not be the omniscient being supposedly proved to exist. That is to say, he would not know himself to be omniscient, and,

therefore, would not be. If that is correct, then there is no omniscient being, and hence no sound argument, ontological or otherwise, for the existence of an omniscient being.[10]

Two Objections

This conclusion is reinforced by consideration of two objections, one pertaining to the relation between omniscience and the logical possibility of error and the other to the modal status of the logical possibility of hiders.

First, I did not mean for my argument to depend upon a contentious conception of knowledge. I claimed that we only need to agree that knowledge is more than true belief and that the 'more' includes non-accidental correctness. My argument might be challenged, though, by questioning whether something more demanding is needed. In particular, it is often supposed that omniscience differs from ordinary knowledge at least in the respect that it logically excludes, not just error, as does any true knowledge claim, but also the logical possibility of error.[11] But then, the objection runs, it might be that I am tacitly importing an assumption to which my uncontentious conception does not entitle me.

Elaboration is in order. In ordinary cases, we agree that if Susan knows something, she must not simply have gotten matters right by accident, but we do not insist that her grasp of the facts be necessary. Susan may know, while being fallible. Her knowledge must exclude error, but need not exclude the logical possibility of error. She can know that it is raining, without being able to rule out the possibility that she is being deceived by a special-effects team dispatched from a movie studio. Her belief may still be non-accidentally true, given that she formed it responsibly and with adequate attention to relevant evidence. Thus, the sense in which her belief is *non-accidentally* correct is not equivalent to its being *necessarily* correct.

This is important to sort out, for I have tried to show that there are facts which no being could know. My argument has been that it would not be possible to distinguish states of affairs in which one of those facts obtained from others in which it did not obtain. Thus, any being is at most only *supposedly*, rather than *actually*, omniscient. But that description suggests that my argument may rest upon the confusion of non-accidental with necessary correctness. To put the objection briefly, can my argument be defused if an omniscient being is non-accidentally right about everything, though possibly mistaken about some things? Could Todd (or God) know every fact, including the non-existence of hiders, without being able to rule out their logical possibility? If so, my argument fails.

In this case, it is the objection that fails. Note first that if Todd cannot rule out the possibility of hiders, then, if Todd is omniscient, they must be possible. Thus, the objection presupposes that hiders are logically possible. But, given that their defining property is that they are capable of perfectly concealing their existence from any other being, they cannot be known, by some other being, either to exist (if they are in hiding) or not to exist (if they do not). They could only be known not to exist if they were in some way impossible. The objection fails because it presupposes the possibility of hiders, but there can only be an omniscient being if hiders are impossible. Whether or not there is room for an omniscient being to know something while being possibly mistaken about it, he cannot be possibly mistaken *about hiders* and know that there are none.

A second objection questions the claim that hiders are logically possible. The core of my argument has turned upon the claim that hiders are possible and therefore that no being is omniscient. Some theists might reply, however, that what I have shown is at most that hiders are *narrowly* logically possible—roughly, that they cannot be shown to be impossible with our actual formal calculi—but have not shown that they are broadly logically possible.[12] Since narrow logical possibility does not entail broad logical possibility, it may be that hiders are broadly logically impossible, even if we do not have the formal logical machinery to prove it. Thus, no conflict has been demonstrated between the possibility of hiders and the actuality of omniscience, for hiders may only be narrowly logically possible, but broadly logically impossible.[13]

I accept this point. I have not shown hiders to be broadly logically possible. Thus, it may be, so far as I have shown, that hiders are broadly logically impossible and thus do not constitute a barrier to the possibility of omniscience. Of course, it might instead be that it is omniscience that is broadly logically impossible. Since the broad logical possibility of omniscience entails the broad logical impossibility of hiders, and the broad logical possibility of hiders entails the broad logical impossibility of omniscience, and since we cannot show which of the respective possibility-impossibility pairs obtains, it might be suggested that the only recourse is to faith.

In this case, recourse to faith is premature. Even if hiders are broadly logically impossible, the objection does not address the identification problem. If we suppose that some Arguer has in hand a proof that there is an omniscient being, and so, knows that hiders are broadly logically impossible, that does not enable the Arguer to identify *herself* as the omniscient being. To draw the conclusion that the Arguer is omniscient, some premise will be needed to the effect that the Arguer possesses some characteristic that only an omniscient being could possess, but that premise, it appears, will be just as doubtful as that the Arguer is omniscient. Even if, say, only an omniscient being could be omnipotent, there seems to be no way for any but an omniscient being to know her

own omnipotence. The argument will be circular, so the Arguer cannot know herself to be omniscient.

Other versions of this objection are equally infected by the identification problem. For suppose that there is some proof that hiders are impossible that does not entail the existence of an omniscient being. Then, two further steps will be necessary. First, the proof that hiders are impossible will at most show that it is possible for there to be an omniscient being, not that there actually is one. Some further reason will be needed for thinking that there actually is an omniscient being. Then, with that reason in place, the Arguer will still need some reason for identifying herself with the omniscient being. For the same reasons as above, the argument will be circular.

In the end, the identification problem undermines the original supposition that there is some proof of the existence of an omniscient being. Since no Arguer can identify herself as omniscient, any supposed proof to the contrary must have been unsound.

Conclusion

Let's take stock. Consider the following argument:

(1) If any being is omniscient, then it knows that it is omniscient.
(2) No being knows that it is omniscient, because there is no being that knows hiders to be impossible and who can solve the identification problem.
(3) Therefore, no being is omniscient.

That is certainly valid, and the first premise seems beyond question, so it is sound if the second premise is also true. Let's rephrase the second premise as the conjunction of (2a) and (2b):

(2a) If hiders are not known to be impossible or if the identification problem is not solved, then no being knows that it is omniscient.
(2b) Hiders are not known to be impossible, or the identification problem is not solved.

(2a) seems to be correct. There are several ways to challenge (2b), but, as I have argued, even if it were possible to adequately address the first disjunct, none of them provides an adequate response to the identification problem, so it appears that (2b) is also true. In my judgment, then, the argument is sound. So we have

(1) If any being is omniscient, then it knows that it is omniscient.
(2a) If hiders are not known to be impossible or if the identification problem is not solved, then no being knows that it is omniscient.
(2b) Hiders are not known to be impossible, or the identification problem is not solved.
(3) Therefore, no being is omniscient.

Thus, we appear to have a sound argument that shows that there is no omniscient being and, further, one that provides a novel line of criticism of the ontological argument.

Though I take it to be sound, I do not claim to be able to see that my premises are necessary truths. Still, though not necessary so far as I can tell, the premises seem uncontentious and obvious. For those who think the conclusion mistaken, the indicated course is to question the truth of the premises. To the extent that I am correct that the premises are uncontentious and obvious, my argument at least raises the stakes. It shows how much has to be given up to reject its conclusion.[14]

Notes

[1] See Bertrand Russell, *The Problems of Philosophy* (New York: Oxford UP, 1978) 46ff.

[2] I shall not be concerned with other conceptions of omniscience. In particular, I exclude Swinburne's conception, according to which God knows all that can be known, but which allows that there may be unknowable truths (e.g., Richard Swinburne, *The Coherence of Theism* [Oxford: Clarendon Pr, 1977]). That conception may have other problems—I suspect it does—but it is immune to the argument I develop here.

[3] Since hiders are defined as able to conceal their existence from all others, the same being might be both omniscient and a hider. For simplicity, I assume that we are speaking of hiders distinct from the supposedly omniscient being.

[4] Since hiders are defined in terms of their ability to conceal themselves from *other* beings, it may be unclear whether there are any hiders in possible worlds with no other cognizing beings. In such worlds, should we say that there are *no* hiders or that *everything* is a hider? Since we are interested in the possibility of omniscience, and therefore in that subset of possible worlds that contain at least one cognizing being, we need not settle this. It represents no loss of generality to exclude from consideration those unclear cases in which there are no cognizing beings from which to hide.

[5] I mean to include here the possibility of disjunctive facts. It may be that there are no hiders in worlds α or β because the disjunctive fact, *p or q*, holds in both, that *p* is incompatible with hiders, that *q* is incompatible with hiders, that *p and not-q* holds in α, and that *q and not-p* holds in β.

[6] I do not mean to be denying reliabilism here, though I have serious doubts about how one might reliably but not conclusively determine that hiders are impossible. That need not be sorted out here, however, since we are addressing only the question of whether Todd might have reliable *non-inferential* knowledge of the impossibility of hiders. As argued in the next paragraph, if hiders are impossible, then that fact must be inferentially knowable as well. There will be some sound argument that hiders are impossible from premises that Todd knows to be true.

[7] I further consider the issues raised by independent proofs of an omniscient being below.

[8] I speak of a modal version of the ontological argument because that, particularly as worked out by Alvin Plantinga in *The Nature of Necessity* (Oxford: Oxford UP, 1974), seems to me the most philosophically interesting. However, the intricacies of Plantinga's version are not essential to my discussion. What is essential for my purposes is that the being supposedly proved to exist is maximally or unsurpassably great, and that such unsurpassable greatness entails propositional knowledge of all facts.

[9] The ontological argument also does not seem to prove that the set of traits definitive of maximal greatness is *uniquely* exemplified. Perhaps, there are several tied at the top. It still might be true of each one that no greater can be conceived.

[10] It might be objected that if omniscience is not a possible property, its non-existence cannot be an objection to the ontological argument, for the most that the argument might be supposed to prove is that there is some being with the maximum possible knowledge. This response supposes that there is a coherent conception of maximum possible knowledge. There may not be, however, any more than there is a largest integer. The ontological argument is usually understood to underwrite the attribution of omniscience to God on the grounds that, in order to be unsurpassably great, God must be noetically unsurpassable, but would not be noetically unsurpassable if his knowledge were limited in any way. His noetic greatness could be surpassed by a being that was not so limited. But it might be that each and every degree of noetic greatness is itself limited and thus surpassable. If so, there will be no such thing as unsurpassable noetic greatness. To illustrate the point in the present context, it may be that there are *relative* hiders—those relative to human faculties, whose existence cannot be detected by human beings. Martians may be less limited and able to detect the human-relative hiders, but unable to detect Martian-relative hiders. Jupiterians may be able to detect Martian-relative hiders, but not Jupiterian-relative hiders. And so on. I would like to thank Renée Smith for encouraging greater clarity on this point.

[11] I take no stand on this question. I suspect that omniscience does not admit the logical possibility of error, but, since the present objection can be addressed without settling that question, I will reserve that argument for some other occasion.

[12] See Plantinga 1-2 for the distinction. As Plantinga characterizes it, narrowly logical necessity comprises the "truths of propositional logic and first order quantification theory."

[13] The objection is stated in a way that involves minimal commitment on the theist's part—only to the proposition that *something* that is broadly logically necessary makes the existence of hiders broadly logically impossible. The most straightforward way for the theist to hold this is to maintain that hiders are broadly logically impossible because of the existence of a maximally great being, one who is maximally perfect—omniscient, omnipotent, perfectly good, etc.—in all possible worlds. See Plantinga Chapter X. That amounts to a commitment to one form of the ontological argument. Since that raises additional questions beyond my purview here, I shall not pursue it.

[14] I would like to thank Gayle Dean, Carter McCain, Renée Smith, and four anonymous referees for insightful comments and discussion.

References

Plantinga, Alvin. *The Nature of Necessity*. Oxford: Oxford UP, 1974.

Russell, Bertrand. *The Problems of Philosophy*. New York: Oxford UP, 1978.

Swinburne, Richard. *The Coherence of Theism*. Oxford: Clarendon Pr, 1977.

Philosophy of Emotion and Ordinary Language

Scott Kimbrough, *Jacksonville University*

What is an emotion? This question has picked up a lot of momentum in the last thirty years or so, as philosophers, psychologists, and neuroscientists offer competing answers. One of the more hotly contested questions in these debates concerns the role of judgment in defining emotion. "Cognitivists" maintain that judgments are essential to any adequate understanding of the emotions.[1] For instance, contemporary cognitivists follow Aristotle's lead in claiming that anger involves the judgment that one has been wronged. Critics of cognitivism dispute the relevance of judgments to emotion. For example, reviving the James-Lange theory according to which emotions are perceptions of bodily states, philosopher Jesse Prinz contends that the physical bases of emotional response are faster and more neurophysiologically basic than those implicated in cognition. Thus, he concludes, judgments are not relevant to the definition of emotion.[2]

I will argue that Prinz is wrong to deny cognitivism. To make the case, I will deploy a distinction developed by Peter Strawson in his well-known discussion of free will: the distinction between the "participant attitude" adopted within normal interpersonal interactions and the "objective attitude" adopted for scientific study.[3] In the process, I hope to clarify the basis for Strawson's philosophical appeals to "ordinary language" or, as it is now misleadingly known in the philosophy of mind (thanks to a rhetorical victory by its philosophical enemies), "folk psychology."

Begin by considering Prinz's critique of cognitivism. He proceeds in two steps: first, he attempts to establish that the debate between cognitivism and its critics must be settled scientifically; second, he refutes cognitivism as a scientific hypothesis. For the sake of argument I am happy to concede the second step of Prinz's argument.[4] It's the first step I will argue is too hasty.

In his effort to establish that science must settle the relevance of judgments to emotions, Prinz presents the only alternative to scientific studies as "philosophical reflection" that appeals to "intuitions" about examples.[5] Such intuitive appeals predictably do not compel agreement on all sides of the debate, leading Prinz to remark:

> If intuitions can clash, as they almost always do, theories of emotion should also avail themselves of other kinds of evidence…. Reflection may reveal more about the person reflecting than about the phenomenon on which she is reflecting. If one wants to explain

> something other than one's own personal beliefs, one should exploit more objective methods. In particular, one should make use of scientific experiments.[6]

Prinz goes on to praise empirical psychologists for their use of statistical methods, concluding that they "may have a methodological advantage over philosophers," whom he discounts as "casual observers."[7]

Prinz's form of argument proves rather more than it should. For example, suppose the subject of debate were a moral issue rather than the nature of emotions. Surely "intuitions" differ about examples. Is what we need at this juncture "other kinds of evidence" or "scientific experiments"? Shall we call in the empirical psychologists to exploit their "methodological advantage"? Clearly not, as scientific methods – however admirably suited for sifting causal hypotheses – do not confer expertise within a moral debate. Moral debate, as Strawson explained in "Freedom and Resentment," requires us to maintain what he calls the "participant attitude."

Strawson introduces the distinction between the participant attitude and the objective attitude in the following passage:

> What I want to contrast is the attitude (or range of attitudes) of involvement or participation in a human relationship, on the one hand, and what might be called the objective attitude (or range of attitudes) to another human being, on the other. Even in the same situation, I must add, they are not altogether *exclusive* of each other; but they are, profoundly, *opposed* to each other. To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment….[8]

In contrast, when relating to another person with the participant attitude, we expect reciprocal goodwill and consideration, and respond with emotions such as resentment, anger, and indignation when we do not receive it. Strawson calls these emotions (and their more positive counterparts such as gratitude and appreciation) "reactive attitudes," and emphasizes their critical importance to normal moral life. Ascriptions of moral responsibility presuppose that their target is capable of the expected reciprocal goodwill; a finding of incapacity, as in the case of the mentally ill, constitutes one good reason to suspend the participant attitude, including the emotions characteristic of it such as anger and moral indignation, in favor of the objective attitude. The desire to study causes scientifically is another reason to adopt the objective attitude. In the course of their studies,

scientists endeavor to suspend resentment, gratitude, moral indignation and other reactive attitudes characteristic of the participant attitude.

Strawson expects that a scientific account of the reactive attitudes themselves is possible. Indeed, Strawson and Prinz both support a broadly Humean meta-ethics according to which emotions under-gird moral practices.[9] However, Strawson's purpose is not to engage in such scientific research, but rather to point out that an objective account of reactive attitudes does not capture their importance for our lives:

> It is one thing to ask about the general causes of these reactive attitudes I have alluded to; it is quite another to ask about the variations to which they are subject, the particular conditions in which they do or do not seem natural or reasonable or appropriate; and it is a third thing to ask what it would be like, what it *is* like, not to suffer them. I am not much concerned with the first question; but I am with the second; and perhaps even more with the third.[10]

Strawson answers that a life without reactive attitudes would be inhuman, irrational, and impoverishing, depriving us of "the general framework of human life."[11]

Now, I don't think Strawson is likely to meet much resistance to this claim. Prinz emphasizes the survival value of emotions in our evolutionary history, noting that anger for example "may be acquired because it helps us cope with challenges from conspecifics."[12] However, I will argue that Strawson's insight cannot be squared with Prinz's rejection of cognitivism.

At this point, a brief summary of Prinz's non-cognitive theory of emotion can no longer be avoided. Prinz calls his view the "embodied appraisal theory," and promotes it as a synthesis of the James-Lange theory and appraisal theories. Emotions, as he understands them, detect states of the body. This is the Jamesian portion of his theory. However, emotions do not *represent* the bodily states they detect. Rather, these bodily states represent "core relational themes" – relations between an organism and its environment that pertain to well-being.[13] This is the appraisal part of the theory. Unlike appraisal theorists such as Lazarus, however, Prinz maintains that it is the bodily states themselves that represent the core relational themes, rather than cognitive judgments or "propositional attitudes" involving the relevant concepts. Fear, for example, is a non-cognitive, non-conceptual bodily state that *represents* danger without involving the *judgment* that danger is present. Of course, the judgment and the bodily state often co-occur. Indeed, the judgment can cause or, as Prinz often says, "trigger," the bodily state. But the bodily state can also occur in the absence of the judgment, as when a snake-phobic person reacts immediately to the perceptual image of a snake.

Ambitiously, Prinz maintains that his theory applies not only to "basic" emotions the content of which is rooted in our evolutionary history, but also to "higher" emotions like jealousy that only cognitively sophisticated and culturally influenced creatures like ourselves are capable of experiencing:

> I propose that cognitively elaborated embodied appraisals are not composite states at all. They are comprised of nothing but embodied appraisals. The cognitions that elaborate them are *prior conditions*, not *constituent parts*. When romantic jealousy occurs, there is first a judgment to the effect that one's lover has been unfaithful and then an embodied appraisal. The emotion, jealousy, is comprised entirely by the embodied appraisal.[14]

On this theory, thoughts about infidelity can trigger jealousy, but jealousy is a non-conceptual bodily state that represents infidelity independently of those judgments.[15]

On the face of it, Prinz's distinction between emotions and triggering judgments is forced, particularly because the existence of the bodily states with which he identifies emotions is, for the vast majority of emotions, speculative. However, Prinz's distinction is admittedly not arbitrary: one main point, as I understand it, is to direct empirical research of the emotions away from higher cognition and towards non-cognitive brain systems. This may well be a more worthwhile direction for empirical research into the physiological aspects of emotional experience. It certainly strikes me as more promising than administering surveys, as many psychologists studying emotions do. Nevertheless, even if Prinz is right about how best to study the brain, it does not follow that he has captured all that matters about emotion. As Strawson's discussion of the participant attitude reminds us, our daily concern with emotions plays a role in our moral practices, not just our scientific ones.

But might the advance of science change those practices? Strawson acknowledges that scientific study can affect our understanding of concepts within the participant perspective.[16] This has obviously happened already, as the medicalization of mental illness has substantially reduced the moral condemnation of people suffering from depression and other affective disorders. Would we all come to speak about emotions like Prinz if scientific research substantiates his theory about the bodily basis of emotion? Would it be pertinent to distinguish, with Prinz, between the emotion strictly so called and the judgments that merely "trigger" the emotion?

For example, suppose that scientific research of the sort favored by Prinz yields a "treatment" for anger – an injection that prevents the onset of anger's typical physiological manifestations.[17] At the same time, suppose that the injection has no effect on judgments, allowing the patient to continue to make the judgments emphasized by cognitivists as necessary to anger (e.g.,

that one has been wronged, that the wrong was inflicted intentionally, etc.). Would the invention of such a treatment prove that anger does not involve judgments? It certainly seems natural enough for an injected patient to report his experience by saying: "I see that he wronged me on purpose, but I just can't get angry about it." But this shows at most that physiological arousal is *necessary* for emotions, a thesis that most cognitivists (excepting notable hold-outs such as Nussbaum) do not dispute.

To determine how or whether judgments are also essential to our understanding of emotions like anger, consider this question: How would we decide whether to prescribe the injection? This question clearly draws on moral, not merely physiological concerns. The injection would have to be reserved for patients who get angry in inappropriate circumstances, and whose subsequent behavior proves destructive to themselves and others. In other words, to judge whether a person's anger is *normal* or *pathological*, we must examine the judgments the person makes when angry.[18] Only a person with a frequent habit of becoming angry for the wrong reasons, and acting on that anger in the wrong ways, would be a candidate for the treatment.[19] Even in a medical context, discussions of anger must draw on judgments of reasonability, appropriateness, proportionality, etc.

Solomon's cognitivist theory emphasizes such judgments. He argues that theories of emotion like Prinz's lead to a morally irresponsible conception of emotions as out of our control:

> What has always centrally concerned me is what Sartre captured in his famous concept of 'bad faith' (*mauvaise foi*), that is, our tendency to deny responsibility by making excuses for ourselves. One of the most prevalent modes of bad faith is our blaming our behavior on our emotions and so excusing ourselves from any responsibility.[20]

Prinz certainly plays into Solomon's concern. Although he concedes that emotions can be evaluated via criticism of the judgments that trigger them, he maintains that emotions themselves cannot be criticized. Commenting on a person whose "calibration file" (i.e., set of triggering causes, including judgments) for amusement includes others' misfortune, Prinz writes:

> It may be wrong, in some sense, to have such a file. But now ask, is it wrong to laugh in the face of adversity assuming such a calibration file is already in place? Put in this way, the answer may be negative. It is not wrong to feel amusement when one encounters something that matches the contents of your amusement file. Nor is it right. Once a calibration file has been set up, we cannot help but react to its contents. This is one source of emotional

> passivity. The response to items in our calibration files is automatic, and falls outside the jurisdiction of normative assessment. Viewed in this light, emotions are really arational.[21]

In Strawsonian terms, what we have here is Prinz insisting that emotions must be evaluated objectively, not from the participant attitude.

To highlight the disagreement between Prinz and Solomon, consider the legal case of *Kansas v. Borman* (1988).[22] Borman killed his girlfriend after she went on a date with another man. He sought to defend himself by appeal to his "intermittent explosive disorder" – an angry personality. Given his personality, the defense argued, Borman could not control his anger when he discovered his girlfriend's infidelity. He admitted to killing her, but claimed "diminished capacity." We can imagine Prinz serving as an expert witness in the case, testifying that Borman's personal history had set up a calibration file for anger and jealousy that automatically generated an overwhelmingly angry and jealous response when his girlfriend was unfaithful.[23] However, the judge instructed the jury to evaluate the diminished capacity claim as follows:

> The criminal law concept of diminished capacity requires the presence of a mental disease or defect not amounting to legal insanity which a jury may consider in determining whether the defendant has the specific intent required for the crime charged. Mere personality characteristics such as poor impulse control, a short temper, frustration, feelings of dependency, "snapping," lack of concern for the rights of other people, etc., do not constitute a mental disease or defect bringing the doctrine of diminished capacity into play.

Despite the testimony establishing Borman's "impulsive and immature" character, the jury concluded that he could not avoid criminal liability: he was convicted of second degree murder (intentional, but unpremeditated murder).[24]

The court's reasoning echoes ordinary examples of moral assessment. When a person is angry, he may sensibly be asked "Why are you so angry?" To this question, a reference to one's "calibration file" constitutes the wrong sort of answer. What's expected is an explanation of what the object of one's anger did to deserve such a reaction. The focus on judgment in these contexts is not a focus on causes of anger, but on reasons for anger. When relating to other people normally from the participant attitude, we expect (in the normatively freighted sense of 'expect') the angry person's anger to be bound up with understandable reasons. If it isn't, we hold them responsible. To retreat to talk of causes at this point is, like Borman, to offer unacceptable excuses.[25] Accordingly, Prinz's effort to shave judgments off of emotions doesn't square with how responsibility is

attributed in actual practice. These practices make much less sense if we follow Prinz in characterizing emotions as non-cognitive, arational states.[26]

None of this is meant as an attack on psychology. Rather, the point is that psychological study is one of the things we do, and that the significance of that research is something that must be interpreted in the light of our other interests, goals, and purposes. When philosophers like Strawson draw our attention to our extra-scientific judgments, they are not typically engaged in "folk psychology," an armchair attempt to explain and predict behavior. Nevertheless, their concern is and always has been truth. But what kind of truth is it?

One answer often floated to this question by both critics and proponents of "folk psychology" draws on a distinction between empirical truths and truths about meaning – truths that any competent speaker comes to know in virtue of learning the language. Griffiths' eliminativist polemic features this view, which he rejects on the grounds that "conceptual analysis" of emotion concepts only tells you what people currently believe about emotions rather than what is true.[27] Ordinary language philosopher Oswald Hanfling encourages this reading as well, frequently adverting to linguistic competence as the source of justification for claims concerning "what we say" about topics like freedom, knowledge, etc. For example, Hanfling insists that platitudes of "folk psychology" are not *empirical* truths within a proto-scientific theory, but *logical* truths known by anyone who has learned the language.[28]

I agree with scientifically minded critics of "folk psychology" that appeals to the meanings of emotion terms are unhelpful. The right way to counter the over-reaching claims on behalf of science is not by declaring everyone an expert in the logic or (even more obscurely) the "grammar" of English. For such a strategy makes it mystifying how disagreements among competent speakers are possible, and thereby encourages philosophers like Prinz to dismiss non-scientific "intuitions" as arbitrary and subjective. I agree with Prinz and Griffiths that "conceptual analysis" does not yield *a priori* truths forever immune to criticism that draws on empirical discoveries: defending claims about the meaning of emotion terms is neither more nor less difficult than defending claims about the emotions themselves. Nevertheless, I do not conclude, as Prinz and Griffiths do, that science can therefore claim sole authority to determine the nature of emotion. Appeals to ordinary practice, including Strawson's appeal to the participant attitude, work by emphasizing contexts in which we do not in fact accept scientific evidence as decisive. When making and defending an ethical judgment, for example, appeals to causal factors are, if not always irrelevant, frequently wide of the point. When an otherwise normal person insults me, I do not care to hear that his calibration file includes contempt for college professors. His rudeness deserves a rebuke despite the causal

provenance of the remark. This is an ethical point, not an effort at armchair psychology or a "grammatical" point backed by my status as a competent speaker.

The value of ordinary language philosophy, as I see it, lies in reminding us of the point of our concepts in their normal uses. As Wittgenstein insists, it's a mistake to suppose that there is a general form of explanation underlying these reminders.[29] There is no epistemological theory about meanings or "folk-psychological" theory about hidden causes lurking in the reminder that we hold people like Borman responsible for their emotions. Because Prinz always seeks causal explanations in his study of emotion, I can imagine him wondering how Strawson's "participant attitude" is to be coordinated with perception, memory, and other psychological mechanisms.[30] But Strawson is not using the term "attitude" to denote a causal mechanism; rather, he is drawing our attention to the fact that we aren't primarily interested in causal mechanisms when discussing emotions in many ordinary contexts.[31] As examples like Borman's remind us, the practices we participate in commit us to drawing non-severable connections between emotions, judgments, and moral evaluation.

In conclusion, I don't know whether cognitivists like Solomon would welcome my associating their work with the ordinary language tradition. Solomon's heroes are Sartre and Nietzsche, not Austin and Wittgenstein. But he does echo Strawson by making the connection between emotion and responsibility central to his work:

> I want to rest my case on practical and moral considerations, and I would urge that these should always be kept in mind. The main consideration is this: how we *think* about our emotions – as something we suffer or as something we 'do' – will deeply affect both our behavior and our understanding of our behavior. In other words, theses about emotions tend to be self-confirming. If one thinks of oneself as the victim of irrational forces, one need not examine the reasons and motives for acting as one does.[32]

As Solomon makes clear, the problem with Prinz's efforts to separate emotions from judgments is that it prevents us from saying things about people and their emotions that we need to say.[33] Borman is responsible for his action: there is no room for the excuse that his anger is "automatic" or "arational," nor accordingly for a theory of emotion that characterizes his anger exclusively in those terms.

Notes

[1] Some cognitivists, such as Robert Solomon in *The Passions* (New York: Doubleday, 1976) and Martha Nussbaum in *Upheavals of Thought* (Cambridge, UK: Cambridge UP, 2001), maintain the strong thesis that emotions are constituted by judgments alone. On these views, certain feelings or physiological states may co-occur with emotions, but are not relevant to their individuation. Solomon has since declared his earlier view too extreme on the grounds that it neglects other aspects of emotions, including their behavioral and physiological aspects (e.g., see Robert Solomon, *Not Passion's Slave* [New York: Oxford UP, 2003], 196). For a prominent cognitivist theory in the empirical psychology literature, see Richard Lazarus, *Emotion and Adaptation* (New York: Oxford UP, 1991).

[2] Jesse J. Prinz, *Gut Reactions: A Perceptual Theory of Emotion* (New York: Oxford UP, 2004). Paul Griffiths argues for an even more radical, eliminativist conclusion in *What Emotions Really Are* (Chicago: Chicago UP, 1997): that the concept of emotions has no good place in empirical psychology, so emotions do not really exist. Although I will focus on Prinz, my subsequent defense of cognitivism also applies to eliminativist views like Griffiths'.

[3] Peter F. Strawson, "Freedom and Resentment," in *Free Will,* ed. Gary Watson (Oxford, UK: Oxford UP, 1982).

[4] Prinz's argument features the claims that scientific cognitivists such as Lazarus lack neuroanatomical support and cannot account for direct physical induction of emotion (e.g., by drugs or manipulation of facial expressions) (Prinz 40). My main reservation about these arguments is Prinz's assumption throughout that the presence of typical physiological symptoms of an emotion constitute the presence of the emotion. Depending on the case, this may or may not be a plausible assumption. Prinz complains, fairly enough, that any experience that feels like fear is fear. But in the absence of any judgment that something is dangerous, is it so clear that the physiological arousal characteristic of fear (assuming there is one such kind of arousal) will *always* feel like fear? For example, Solomon (*Not Passion's Slave*, 221) notes that physiological reactions can last longer than the corresponding emotions. Commenting on the example of a close encounter with a rock-slide, he points out that the "sweating and palpitations" can continue after the fear is gone. In this case, the same physiological reactions that accompanied the fear during the episode are experienced differently once the danger subsides. It might turn out that, in all such examples, the physiological mechanisms in play differ before and after we say the fear has gone. But I don't think we have to wait for the results of further empirical study to accept the accuracy of Solomon's description of the case.

[5] Prinz 28.

[6] Prinz 29.

[7] Prinz 30.

[8] Strawson 66

[9] Strawson 79. Prinz is working on a book to be entitled *The Emotional Construction of Morals*.

[10] Strawson 64.

[11] Strawson 70.

[12] Prinz 66. Similarly, on empirical grounds, neuroscientist Antonio Damasio argues, in *Looking for Spinoza* (New York: Harvest Books, 2003, 140f), that "social emotions" such as embarrassment are essential to rational decision making: people with lesions to the parts of the brain controlling social emotions cannot manage mundane matters such as financial affairs despite the fact that their higher cognitive abilities (mathematical ability, etc.) remain intact.

[13] Prinz 16.

[14] Prinz 98-99.

[15] Prinz maintains that "From a brain's-eye point of view, basic and nonbasic emotions are alike" (Prinz 158). Nevertheless, Prinz's strategy for defining "higher" emotions is to identify a basic set and then derive others from processes of blending and calibration. Shame, for example, is "a sadness that has been calibrated to one's own violations of a moral code" (Prinz 148). The project here is in some ways analogous to René Descartes' effort, in *Passions of the Soul*, in *Selected Philosophical Writings of René Descartes*, ed. & trans. by J. Cottingham et. al. (Cambridge, UK: Cambridge UP, 1649/1988), to define all emotions as mixtures of joy, sadness, desire, love, hate, and wonder.

[16] Strawson's point echoes J. L. Austin, the paradigmatic ordinary language philosopher. Austin acknowledged that our ordinary understanding of concepts can be improved by scientific study. See Austin's "A Plea for Excuses" in *Philosophical Papers*, third edition, ed. J. O. Urmson and G. J. Warnock (Oxford, UK: Oxford UP, 1961), 189. Austin also advises study of examples from the law (Austin 187), a recommendation I will adopt later.

[17] This exercise is in the spirit of Strawson's question, quoted above, of "what it would be like, what it *is* like, not to suffer [reactive attitudes]."

[18] Note the difference here between anger and pain (and other medical conditions the definition of which does not involve judgment). Ethical deliberation does enter into treatment decisions regarding pain, as when one must decide if lessening one's pain justifies the risk of addiction to the medication. But there is no need to assess the *appropriateness* of pain, as there is for anger. Cognitivism earns its keep with respect to those judgments.

[19] Of course, a person who received the treatment would not exactly be restored to health. The anger-elimination treatment is an amputation: it eliminates the pathology, but not by making the patient whole. The patient could in principle be "cured" if the dosage of the injection could be modulated so that the physiological arousal necessary for anger were reduced rather than eliminated. What would count as a cure could only be determined by judging the patient's ability to get angry *appropriately* – when, and only when, he is able to restrict his anger to situations in which he reasonably judges that he has been wronged, that the wrong was inflicted on purpose, etc.

[20] Solomon, *Not Passion's Slave*, 198.

[21] Prinz 240.

[22] The text of the decision is available online at:
http://www.kscourts.org/kscases/supct/1998/19980417/76131.htm.

[23] Recall the passage quoted above in which Prinz claims that jealousy can be triggered by judgments about infidelity. Prinz does have a strategy for maintaining that Borman's action is blameable. He can say that the judgments that caused Borman's anger were unfounded, or emphasize Borman's ability to restrain his behavioral response to his anger. My question for Prinz at this point, however, is this: If the objective attitude is the appropriate attitude to take to Borman's anger, why not the rest of his psychological states? After all, a completed cognitive psychology will presumably treat cognition and behavioral response in no less causal terms than emotional reactions. To insist on the objective attitude with respect to emotion, while allowing the participant attitude with respect to judgment, is not sustainable. In any event, the case of *Kansas v. Borman* shows that ordinary moral evaluation from the participant perspective does not follow Prinz in withdrawing emotions from normative evaluation.

[24] Borman's sudden anger does help excuse him from the charge of first degree murder – a crime which requires premeditation; but it does not release him from criminal liability altogether, nor even reduce his charge to manslaughter. The "heat of passion" may sometimes reduce the charge to manslaughter if the emotions and actions involved are relatively reasonable. For example, a man who fatally punches another man who deeply insults him is more likely to be convicted of manslaughter than murder. If the fatal blow is administered with a tire-iron, however, a murder conviction becomes more likely. And if the insulted man hunts down and fatally shoots his victim after his anger simmers for a week, a murder conviction is all but assured. Obviously, the test here is not the presence or absence of strong emotion, as the defendant is enraged in all three scenarios (as was Borman). The question juries must decide is not whether the defendant acted in anger, but the

reasonableness (or lack thereof) of both the anger and the actions motivated by it. My thanks to lawyers David and Patty Barksdale and Brian Foley for helping me sort out the legal issues.

[25] Causal answers to the question why one has a particular belief are often out of place as well. Oswald Hanfling provides an instructive example in his book, *Philosophy and Ordinary Language* (London: Routledge, 2000), 257:

> Asked why S believes that the thoughts of Chairman Mao can enable us to surpass all obstacles, one might give an empirical explanation: 'He was brought up and indoctrinated in the People's Republic of China.' (It is an empirical fact that such believes can be caused in such a way.) But usually this is not what is wanted – especially when the question is addressed to the believer. Asked why he believes that p, he would probably be expected to *justify* the belief, as one that is likely to be *true*.

Notice that Hanfling emphasizes the context in which one is actually talking politics with the person who reveres Chairman Mao. This is the participant perspective emphasized by Strawson. Hanfling's example involving belief illustrates Strawson's point that the objective attitude is opposed to the participant attitude.

[26] The relevance of judgment to emotion becomes even more apparent when socially sophisticated emotions are considered. For example, take contempt, which presupposes nuanced judgments about one's position in social hierarchies. In his book *The Anatomy of Disgust* (Cambridge, MA: Harvard UP, 1997), 207ff., William Ian Miller distinguishes various styles of contempt that are at home in different social organizations, including what he calls "upward contempt" (i.e., the contempt of social superiors by social inferiors). Even upward contempt admits of distinct species, depending on the social order. The contempt of the aristocrat's servant for his master differs from the contempt of the brick mason for the "unmanly" law professor (to use Miller's own self-deprecating example) in a contemporary democracy. Miller's deeply insightful discussion does not observe Prinz's distinction between emotions and triggering judgments, and I submit it is a richer account as a result.

[27] Griffiths 5. Prinz's rejection of "intuitions" is similar.

[28] Hanfling 251-2. Griffiths puts the point the same way (23), though of course he disagrees with it. Hanfling's distinction echoes some figures from the ordinary language philosophy tradition. For example, in *Must We Mean What We* Say? (Cambridge: Cambridge UP, 1969), 4 and 13, Stanley Cavell writes that native speakers of English do not need evidence for their claims about "what we say," as

if such claims were sloppy efforts at field lexicography. According to Cavell, being a native speaker is all the qualification one needs to make such pronouncements.

[29] Ludwig Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe (New York: Macmillan Publishing, 1958), §109.

[30] Solomon provides another puzzler for those seeking causal theories when he states that emotions are "subjective engagements in the world" (Solomon, "Emotions, Thoughts, and Feelings," 77). To ask how this terminology fits *within* an empirical psychological theory is to miss its point.

[31] Griffiths acknowledges that the ordinary use of emotion terms is not purely descriptive, but "contains prescriptions about how people should behave, and collective pretenses about how people do behave" (Griffiths 10-11). He goes on to say that "These *mythical aspects* of emotion cannot be smoothly integrated into a developing scientific understanding of human psychology" (emphasis added). Strawson anticipates such reactions. Griffiths' eliminativism about emotion is similar to scientifically motivated incompatibilism about free will, to which Strawson responds as follows:

> [The] prestige [of science] is great, and is apt to make us forget that in philosophy, though it also is a theoretical study, we have to take account of the facts in *all* their bearings; we are not to suppose that we are required, or permitted, as philosophers, to regard ourselves, as human beings, as detached from the attitudes which, as scientists, we study with detachment. This is in no way to deny the possibility and desirability of redirection and modification of our human attitudes in the light of these studies. But we may reasonably think it unlikely that our progressively greater understanding of certain aspects of ourselves will lead to the total disappearance of those aspects (Strawson 80).

[32] Solomon, *Not Passion's Slave*, 232. Nussbaum strikes a similar note in favor of her cognitive theory:

> The cognitive/evaluative view implies that emotional content is itself part of a creature's pursuit of flourishing. Given the fact that human beings deliberate ethically about how to live, it implies that emotions are part and parcel of ethical deliberation. If we see emotions as impulses, we will think that we can educate or change them only by suppression…Indeed, a great advantage of a cognitive/evaluative view of emotion is that it shows us where societies and individuals have the freedom to make improvements. If we recognize the element of evaluation *in the emotions*, we also see that they can themselves be *evaluated* – and in some ways altered, if they fail to survive criticism (Nussbaum 172-3).

[33] A similar point can be made about eliminativists like Griffiths. Eliminativism about emotions prevents us from taking seriously many of the things we need to say – like that irritability is a character flaw, that people have a responsibility to control their anger, etc. Advances in science will not obviate the point of these assertions, even if irritability and anger are not the terms in which finished science is couched.

Works Cited

Austin, J. L. "A Plea for Excuses." In *Philosophical Papers*, Third Edition. Ed. J. O. Urmson and G. J. Warnock. Oxford: Oxford UP, 1961.

Cavell, Stanley. "Must We Mean What We Say?" In *Must We Mean What We Say?: A Book of Essays*. Cambridge: Cambridge UP, 1969.

Damasio, Antonio. *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. New York: Harvest Books, 2003.

Descartes, René. *The Passions of the Soul*. In *Selected Philosophical Writings of René Descartes*. Ed. & trans. John Cottingham et. al. Cambridge, UK: Cambridge UP, 1649/1988.

Griffiths, Paul E. *What Emotions Really Are*. Chicago: U of Chicago P, 1997.

Hanfling, Oswald. *Philosophy and Ordinary Language: The Bent and Genius of our Tongue*. London: Routledge, 2000.

Lazarus, Richard S. *Emotion and Adaptation*. New York: Oxford UP, 1991.

Miller, William I. *The Anatomy of Disgust*. Cambridge, MA: Harvard UP, 1997.

Nussbaum, Martha. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge, UK: Cambridge UP, 2001.

Prinz, Jesse J. *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford UP, 2004.

Solomon, Robert C. "Emotions, Thoughts, and Feelings: Emotions as Engagements with the World." In *Thinking About Feeling*. Ed. R. C. Solomon. New York: Oxford UP, 2004.

Solomon, Robert C. *Not Passion's Slave: Emotions and Choice*. New York: Oxford UP, 2003.

Solomon, Robert C. *The Passions.* New York: Doubleday, 1976.

Strawson, Peter F. "Freedom and Resentment." In *Free Will.* Ed. Gary Watson. Oxford, UK: Oxford UP, 1982.

Wittgenstein, Ludwig. *Philosophical Investigations,* Third Edition. Trans. G. E. M. Anscombe. New York: Macmillan Publishing Company, 1958.

Property Rights of Fetuses: Ontological Problems in the Deprivation Argument against Abortion

Jesse Unruh

Of the many and varied arguments maintaining that abortion is morally on keel with murdering an adult human, the deprivation argument against abortion is the most deserving of philosophical attention. The most noteworthy aspect of the argument is its attempt to prove the immorality of abortion while avoiding the issue of personhood altogether. The essential claim of the argument is that aborting a fetus[1] deprives it of its future which is valuable, and since depriving the fetus of something valuable is immoral, then abortion is immoral. In this essay I argue that the deprivation argument against abortion ('deprivation argument') cannot achieve its aim as it stands. Ultimately, the argument is question begging; the argument assumes that it is immoral to deprive a fetus of something valuable, but gives us no reason to suspect that a fetus is the type of thing from which it is wrong to take valuable things.

The form of this piece shall be relatively straightforward; first, I shall give a brief account of the deprivation argument and some of the specific versions. Then we shall see how all of these formulations share the same fault of assuming their own conclusion.

I. The Deprivation Argument against Abortion

The deprivation argument is simple on first glance. It roughly goes as follows:

P1 – A fetus has a future that is valuable.
P2 – An abortion deprives that fetus of its future.
P3 – Depriving a fetus of something valuable is immoral.
C – Abortion is immoral.

This appears to be a plausible account of the immorality of abortion—if abortion is immoral—but the argument needs some clarification if we are to grasp the different ways in which the deprivation argument can proceed.

I.i. Marquis, Stone, and Feldman

The three versions of the deprivation argument that I am using in this piece are put forth by Don Marquis,[2] Jim Stone,[3] and Fred Feldman.[4] Each of these pieces puts forth a different version of the deprivation argument; however, I will show that they all share the same question begging nature.

Marquis asserts that the reason abortion is wrong is the same reason it is wrong to kill you or me. On Marquis's account, killing is wrong because of the victim's loss of life, which is (presumably) one of the greatest losses one can suffer.[5] The loss consists not only of the total value one's life holds now, but the total value that one would have had in the future were one not killed.[6] Since a fetus typically has a future outside of the womb which will contain value (or in Marquis's terms a "future like ours") the same line of reasoning that applies to the adult applies to the fetus. Marquis therefore concludes that the "correct ethic of the wrongness of killing can be extended to fetal life and used to show that there is a strong presumption that any abortion is morally impermissible."[7] It should be mentioned that there has been an effective critique of Marquis's claims based on an equivocation of the term 'loss' throughout his piece put forth by Walter Sinnott-Armstrong.[8] However, this is not a fault shared throughout all deprivation arguments so it will not be included in the subsequent discussion.

Stone's account is generally similar to Marquis's account. However, Stone is focused on the actualization of the developmental path which will likely lead to the actualization of consciousness which is a great good for the fetus. For Stone, "If the biological nature of an infant is to make herself self aware, then it is in her interest to grow up."[9] Since the nature of the infant and the nature of the fetus (with regards to development) are the same, then the fetus has an interest in growing up just like the infant. And a fetus which would follow its genetic path to its good end is harmed by death.[10] Moreover, we owe it to the fetus "to look out for their interests."[11] In Stone's account it is depriving a fetus of its interests that is wrong.

Feldman's argument is relevantly similar to Marquis's and Stone's, though Feldman concludes that not all abortions are immoral. The morally relevant category for Feldman is desert. When an abortion is immoral (generally late term abortions on his account) it is because we are depriving the fetus of what it deserves, which is its life. The cause for this desert is the fetus's investment in its life that consists of growth and development. So a fetus which has put little to no investment in its life (i.e., an early term fetus) deserves relatively little, and if an eight month fetus is aborted then "her efforts are deprived of their point, and she suffers a serious injustice."[12] (For Feldman, the mother has a certain level of desert that can outweigh the desert of the fetus for a time, which is why a later abortion is immoral but an earlier one is not.)[13]

These three positions can be roughly summarized by the outline of the deprivation argument provided above. One must simply understand that the thing of value for each argument is a "future like ours" for Marquis, interests in growing up for Stone, and desert for Feldman. Moreover, it should be recognized that the moral wrong that occurs with abortion in the above accounts is equivalent with the type of moral wrong that occurs when an adult human is killed.

II. The Ontology of Fetuses[14]

One of the most fundamental challenges facing the moral status of abortion is positing what the fetus essentially is in a moral sense. It seems as though a sufficient answer to this question would lay many of the issues about the morality of abortion to rest. H. Tristram Engelhardt Jr. addresses this issue directly in his essay "The Ontology of Abortion." In setting out the issue he says this about the ontology of the fetus:

> Thus, by ontological status I shall mean certain general categories of being, such as being an inanimate object, being a mere animal, being a fully developed self-conscious human person. With regards to the question of abortion, this is the issue of whether the fetus shows itself to be something to which one owes obligations in the sense one owes obligations to persons.[15]

Marquis certainly agrees with Engelhardt as well as others[16] that the issue turns on what kind of a being the fetus is. At the outset of Marquis's piece he says, "…whether or not abortion is morally permissible stands or falls on whether or not a fetus is the sort of being whose life it is seriously wrong to end."[17] Marquis, as we have seen, goes on to assert that since a fetus has a future of value and depriving someone of a future of value is the central "wrong-making" quality of killing, abortion is wrong. Stone certainly agrees that the abortion issue is dependant on the ontological status of the fetus. His account states that:

> The fact that an infant has the strong potentiality of becoming an adult human being is the fact that if she follows to the end the developmental path primarily determined by her nature which leads to the adult stage of members of her kind, the infant will produce an adult human animal that was once the infant. This fact, conjoined with the fact that adult human animals are self aware, entails that the infant has a nature, the actualization of which involves a great conscious good for the infant. This last fact, I have argued, grounds an interest in growing up.[18]

In other words, abortion is immoral because the fetus is the sort of being that has a grounded interest in growing up just like the infant. And of course Feldman's account asserts that the late-term fetus is the sort of being that deserves not to be aborted.[19]

II.i. Deprivation in Terms of Ontology

Given these accounts, we can generalize premises for the deprivation argument put in terms of ontological status:

P1 – A fetus is the sort of being that has a future and it is valuable.
P2 – An abortion deprives that fetus of its future.
P3 – A fetus is the sort of being that it is immoral to deprive of something valuable.

C – Abortion is immoral.

This way of putting forth the argument seems to better capture the view of the ontological status of the fetus. However, I do not think this account is tenable.

I deny the truth of the above argument due to P3. That is to say that I do not think that the deprivation argument itself has given any good reasons to think that fetuses are the sorts of beings which it is immoral to deprive of valuable things; it has only given us reason to think that a fetus *has* something valuable that can be taken from it.

The grounds on which I base this claim are that of property rights. One of the essential characteristics of any situation where there has been a wrongful deprivation is that whoever it was that was wrongfully deprived of something must have had a right to have it. It is the violating of this right that makes the deprivation a *moral* infraction. Put another way, someone who has been robbed has been the victim of a moral violation only in virtue of the fact that it was her belongings which were taken *and* she is the sort of being from whom it is wrong to take belongings. So any citation of a moral wrong that is based on the deprivation must invoke some notion of property rights.

I should note that while there may sometimes be grounds to differentiate rights and moral obligations[20] I simply mean that the right not to have valuable things taken from an agent and the moral obligation of others not to take valuable things from an agent are the same. This seems to be in line with the literature supporting the deprivation argument. Moreover, if the deprivation argument is to work then both of the above (the right and the moral obligation) would have to be

accepted. In my critique of the deprivation argument I am not trying to separate the two issues and deny one of them, I am simply denying both.

III. To What Do We Have Moral Obligations, and Why?

What is at stake here is, as Raanan Gillon puts it, why it is wrong to shoot peasants while it is ok to shoot pheasants.[21] In this quip, Gillon lays out the problem, that is, to what do we have moral obligations, and why? It is generally agreed that we are obliged not to kill other humans most of the time. This obligation, as Gillon sees it, stems from the nature of other humans; that is to say their human nature.[22] However, just to say that human nature accounts for this obligation does not do much work because almost all philosophically interesting cases are ones that we need to know what counts as 'human nature' or what part of human nature does the work in question. Gillon goes on to give four different possible justifications for the obligation not to kill a being. While there is a useful exploration here, I am not concerned with the usual suspects[23] directly, as they are not argued for by the proponents of the deprivation argument. I am concerned with the deprivation argument, which tries to ground moral agency in having a future of value.

The property of simply having a future of value is not sufficient to establish that a being can be wronged. Clearly, pheasants have a future of value, and even microbes have futures that are of some value. Moreover, most of us are rather sure that the pheasant, or at the very least the microbe, is not the sort of being from which it is wrong to take a future of value. Or at the very least, if there is a moral wrong that occurs when the microbe is robbed of its life, it is not the same sort of wrong that occurs when an adult human is robbed of her life.

So, why are we to suspect that a fetus can be wronged while a microbe cannot? Perhaps it is the type or degree of value that grounds this differentiation. Marquis's piece gives a good account of how we might parse these values; the value most other humans experience from living out a long life might confer moral agency. So it is having value that is derived from having "a future like ours"[24] that differentiates the fetus (and the peasant) from the microbe. However, this account does not work as it stands.

Every cell in our body (except gametes) has a full set of chromosomes that could, under the right conditions, lead to a full human that would have a future like ours.[25] So it must be more than the sheer possibility of a future like ours which grounds moral agency. It seems we also need to incorporate some of Stone's account which stipulates that the differentiation is there because the fetus is very likely to become a person, provided it is not aborted.[26] So, there is a substantive difference in the probability of a certain type of value that a fetus has versus the probability of that

type of value which a human skin cell has, and the type of value a microbe has. This is because a fetus has a high chance of having a future like ours, while a skin cell has little chance of this and a microbe definitely has no such future in store.

However, as special as this value may be, there is still no reason to think that possession of it confers moral agency. While the fetus and persons both possess this unique value the wrongness of killing the person does not derive from the possession of this value alone, it derives from the value itself and the type of being the person is, a moral being. And the reason it is wrong to take things from moral beings is because they are moral beings, not because of what they have which can be taken. The only thing the deprivation argument has established is that *the fetus has something that is valuable which can be taken away*, not that fetuses are the types of beings that can be morally wronged.

The presumption on the part of the proponent of the deprivation argument is that the property 'having a likely future like ours' is what confers the capacity to be wronged. However, this is an unsupported premise. I think that if the obligation for other humans not to kill me is based on a deprivation of my future of value this is only so in virtue of the fact that I am the type of being from whom it is wrong to take valuable things away. We are without cause to think that the fetus is this type of being as well.

The deprivation argument has given us no cause to suspect that a fetus is the type of being from which it is wrong to take valuable things; however, it has not given any reason to doubt that a fetus is that type of being either. Therefore, the question arises whether there are grounds for the proponent of the deprivation argument to claim that a fetus could have property rights. I assert that there is no area for proponents of the deprivation argument to ground their claim based on their premises alone; their argument is incomplete. They must establish that fetuses have rights to things they possess. That is, they must establish whether fetuses are like us in that it is wrong of someone to take something from us under normal circumstances, or if they are more like a microbe, which is not morally wronged when something is taken from it. In other words, do fetuses have property rights?

IV. Fetuses and Property Rights

Many of the central themes of all property rights discussions focus on people or groups of people (like governments or corporations). However, the question we must ask in order to sort out the tenability of the deprivation argument is what sort of beings are eligible for property rights, and moreover, what actually confers those rights on them? If a fetus is the sort of being that qualifies for

property rights then it seems as though the deprivation argument has ground on which to stand, if not, then it seems it is a groundless claim.

One possible, and rather intuitive, answer to the above question is to say that being a person is sufficient to confer property rights. Yet one would have to presume that a fetus was a person for the deprivation argument to be effective, which is immediately problematic. The proponent of the deprivation argument is trying to make a claim about the ontological status of the fetus without simply stating that the fetus is a person.[27] Another intuitive response may be that being *Homo sapiens* is what confers property rights. And since fetuses are *Homo sapiens*, they are the sorts of beings that have property rights. However, the rather obvious problem with this response is that this would disallow property rights for aliens, human genetic mutants, and artificial intelligences. Perhaps *Homo sapiens* is a sufficient condition for having property rights, but not a necessary condition. This would allow for property rights to be conferred upon non-*Homo sapiens*, which is something a good account of property rights should allow for. However, we would also have to include brain dead *Homo sapiens*, which seems contentious in light of our acceptance of organ donation in the case of brain dead patients. Of course we could amend the requirements to be non-brain dead *Homo Sapiens*; however, this caveat does not help the deprivation argument prove that all abortions are immoral. For there is a time frame in which fetuses do not have brains, and thus those fetuses would not have the right to their lives. In addition, anencephalic fetuses would not be able to suffer moral violation. But taking the argument in this direction is missing the point, as the above claims are out of the scope of the deprivation argument against abortion. To take the argument in this direction is to begin to make claims about the bounds of moral agency; claims which are far beyond the bounds of the simple premises put forth in the deprivation argument.

Even if a proponent of the deprivation argument attempted to travel down this path and claim that fetuses do have property rights, it seems they have a lot of work to do to establish this manner of right for fetuses. A brief look at different accounts of property rights will illustrate the hurdles a proponent of the deprivation argument would face.

Classically, property rights are conferred only on moral agents. Hegelian property rights derive from one's moral development in the community[28] while Lockean property rights are acquired through the interaction of the person in the community.[29] However, both of these accounts assume the moral agency of the right bearer. In fact, any type of rights-based property theory would have to assume the moral agency of the right bearer; for if a thing bears rights it is a moral agent. On the other end of the spectrum is utilitarian property rights; however, these assume the moral agency of the individual. That is to say, the utility of the agent is not counted in utilitarian calculus unless the agent is a moral agent. Moral agency is established prior to the discussion of the agent being

eligible for property rights. So, to prove that fetuses are eligible for property rights would entail proving the moral agency of the fetus.

I suggest this difficulty in trying to ascertain how fetuses might come to have property rights is due to the nature of the question being asked. We are asking what manners of beings are entitled to property rights. However, the entire question of which beings have rights is contingent on which beings have moral agency and why they have it. So, to simply say that a fetus has something of value, and that we take it away when it is aborted, says nothing about its ontological status as a moral agent. More directly, the deprivation argument has never given us grounds to think that fetuses are moral agents.

IV.i. Marquis, Stone, and Feldman Revisited

If we revisit the three versions of the deprivation argument presented earlier, we will now be able to see plainly the question begging nature of the accounts. Marquis was concerned with a fetus being deprived of a future like ours; however, the simple fact of having a future like ours does not confer moral agency on an individual. This seems to entail that at some point in the future the fetus will have moral agency, but it does not show the existence of it now. Stone's argument suggests that growing up is in the interests of the fetus. Again, we are faced with the lack of moral agency; just because something has interests does not mean that we are obliged to protect those interests. Microbes have interests, yet we are without fault when we deprive the microbe of those interests. And finally, Feldman's account says that the fetus's investment in life ought not to be taken from it. However, there are many living things which invest time and energy in their lives and for which we are not at fault for taking that investment. To reiterate, the possession of these valuable things does not entail moral agency. And the deprivation argument only works if fetuses are the types of beings that can be wronged.

V. Conclusion

We have seen that the deprivation argument has given no reason to think that fetuses are the sort of beings that have property rights. Moreover, we have seen that an argument against abortion based on a fetus's rights must presuppose that fetuses are moral agents. This is because the deprivation argument must invoke property rights, and to be a candidate for property rights generally, one must be a moral agent. The proponents of the deprivation argument seem to assume that fetuses are relevantly similar to people in more than just the sense that they have futures that

will likely be valuable; they also assume that they are moral agents. Since the moral agency of fetuses is essentially what the argument is trying to prove the deprivation argument does not achieve its goal; it is question begging. Thus, if abortion is morally equivalent to killing an adult human, it is surely not so based solely on the deprivation argument.

Notes

[1] I want to make clear that by 'fetus' I mean the organism from the formation of a single cell zygote to birth. I, like Warren Quinn in "Abortion: Identity and Loss," *Philosophy and Public Affairs* 13.1 (1984): 24, footnote 2, recognize that it is a common (and perhaps bad) habit within philosophical discourse about abortion to refer to the organism by such a general term. However, I will not be basing any moral distinctions on the differentiation between a zygote and a fetus.

[2] Marquis, "Why Abortion is Immoral."

[3] Stone, "Why Potentiality Still Matters."

[4] Feldman, Chapter 12.

[5] Marquis, "Why Abortion is Immoral," 51.

[6] Marquis, "Why Abortion is Immoral," 50.

[7] Marquis, "Why Abortion is Immoral," 55.

[8] Sinnot-Armstrong, "You Can't Lose What You Ain't Never Had: A Reply to Marquis on Abortion."

[9] Stone, "Why Potentiality Still Matters," 2.

[10] Stone, "Why Potentiality Still Matters," 2.

[11] Stone, "Why Potentiality Still Matters," 17, footnote 6.

[12] Feldman 203.

[13] Feldman 204.

[14] In this section it may seem as though I am trying to imply that the moral category "rights-bearer" is some essential or real category of being. I am not claiming this.

[15] H. Tristram Engelhardt Jr., "The Ontology of Abortion," *Ethics* 84.3 (1974): 218.

[16] In addition to Engelhardt, Marquis asserts the following authors share this claim: Joel Feinberg in "Abortion," in *Matters of Life and Death: New Introductory Essays in Moral Philosophy* (1986), Michael Tooley, "Abortion and Infanticide," *Philosophy and Public Affairs* II.1 (1972), Mary Anne Warren, "On the Moral and Legal Status of Abortion," *The Monist* VII.1 (1973), L. W. Sumner, *Abortion and Moral Theory* (1981), John T. Noonan, Jr., "An Almost Absolute Value in History," in *The Moarality of Abortion: Legal and Historical Perspectives* (1970), and Philip Devine, *The Ethics of Homicide* (1978).

[17] Marquis, "Why Abortion is Immoral," 46.

[18] Stone, "Why Potentiality Matters," *Canadian Journal of Philosophy* 17 (1987), cited in Stone, "Why Potentiality Still Matters," 2.

[19] Feldman 203.

[20] Jeremy Waldron, *The Right to Private Property*, (New York: Oxford UP, 1988) 70.

[21] Raanan Gillon, "To What Do We Have Moral Obligations and Why?" In *Contemporary Issues in Bioethics*, edited by Tom L. Beauchamp & LeRoy Walters, 3rd ed. (Belmont, California: Wadsworth Publishing Company, 1989) 161.

[22] Gillon 161-162.

[23] I.e., personhood, sentience, *Homo sapiens*, and viability.

[24] Marquis, "Why Abortion is Immoral," 51.

[25] C. Cameron and R. Williamson, "In the World of Dolly, When Does a Human Embryo Acquire Respect?" *Journal of Medical Ethics* 31 (2005): 215.

[26] Stone, "Why Potentiality Still Matters," 2.

[27] The deprivation argument for Marquis, Stone, and Feldman is proposed as an alternative for personhood arguments against abortion. See Marquis, "Why Abortion is Immoral," 46-49; Stone, "Why Potentiality Still Matters," 2; and Feldman, 201.

[28] Georg W. F. Hegel, *The Philosophy of Right* (Digireads.com, 2004), § 36.

[29] Locke, *Two Treatises of Government*, Chapter 11, Section 137.

Bibliography

Brown, Mark T. "A Future Like Ours Revisited." *Journal of Medical Ethics* 28 (2002): 192-195.

Brown, Mark T. "Abortion and the Value of the Future. A Reply to a Defense of the Potential Future of Value Theory." *Journal of Medical Ethics* 28 (2002): 202.

Brown, Mark T. "The Morality of Abortion and the Deprivation of Futures." *Journal of Medical Ethics* 26 (2000): 103-107.

Cameron, C. and R. Williamson. "In the World of Dolly, When Does a Human Embryo Acquire Respect?" *Journal of Medical Ethics* 31 (2005): 215-220.

Dunstan, G.R. and Marry J. Seller, eds. *The Status of the Human Embryo*. London, UK: Oxford UP, 1988.

Engelhardt Jr., H. Tristram. *The Foundations of Bioethics*. New York: Oxford UP, 1986.

Engelhardt Jr., H. Tristram. "The Ontology of Abortion." *Ethics* 84.3 (1974): 217-234.

Feldman, Fred. *Confrontations with the Reaper*. New York: Oxford UP, 1992.

Foot, Philippa. *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Clarendon Pr, 2002.

Gillon, Raanan. "To What Do We Have Moral Obligations and Why?" In *Contemporary Issues in Bioethics*. 3rd ed. Eds. Tom L. Beauchamp & LeRoy Walters. Belmont, California: Wadsworth Publishing Company, 1989.

Glover, Jonathan. *Causing Death and Saving Lives*. Harmondsworth, Middlesex, England: Penguin, 1997.

Harman, Elizabeth. "Creation Ethics: The Moral Status of Early Fetuses and the Ethics of Abortion." *Philosophy of Public Affairs* 28.3 (1999): 310-324.

Hegel, Georg W. F. *The Philosophy of Right*. Digireads.com, 2004.

Locke, John. *Two Treatises of Government*. Kessinger digital reprint. Indianapolis and Cambridge: Hackett Publishing Company, 1980.

Marquis, Don. "A Defense of the Potential Future of Value Theory." *Journal of Medical Ethics* 28 (2002): 198-201.

Marquis, Don. "Deprivations, Futures and the Wrongness of Killing." *Journal of Medical Ethics* 27 (2001): 363-369.

Marquis, Don. "Korcz's Objection to the Future-of-Value Argument." *Journal of Social Philosophy*. 35.1 (2004): 56-65.

Marquis, Don. "Savulescu's Objection to the Future of Value Argument." *Journal of Medical Ethics* 31 (2005): 119-122.

Marquis, Don. "Why Abortion is Immoral." In *Bioethics an Anthology*. Eds. Helga Kuhse and Peter Singer. Oxford, UK: Blackwell Publishing, 1999.

McMahan, Jeff. *The Ethics of Killing Problems at the Margins of Life*. Oxford: Oxford UP, 2002.

Parel, Anthony and Thomas Flanagan, eds. *Theories of Property, Aristotle to the Present*. Waterloo, Ontario, Canada: Wilfrid Laurier UP, 1997.

Quinn, Warren. "Abortion: Identity and Loss." *Philosophy and Public Affairs* 13.1 (1984): 24-54.

Rosenberg, Jay F. *Thinking Clearly About Death*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1983.

Savulescu, J. "Abortion, Embryo Destruction and the Future of Value Argument." *Journal of Medical Ethics* 28 (2002): 133-135.

Scott, Rosamund. *Rights, Duties and the Body*. Oxford & Portland, Oregon: Hart Publishing, 2002.

Sinnott-Armstrong, Walter. "You Can't Lose What You Ain't Never Had: A Reply to Marquis on Abortion." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*. 96.1 (1999): 59-72.

Stanton, Catherine and John Harris. "The Moral Status of the Embryo Post-Dolly." *Journal of Medical Ethics* 31 (2005): 221-225.

Stone, Jim. "Why Potentiality Still Matters." *Canadian Journal of Philosophy* 24.2 (1994): 281-293.

Stretton, Dean. "The Deprivation Argument against Abortion." *Bioethics* 2.2 (2004): 144-180.

Thomson, Judith Jarvis. "A Defense of Abortion." In *Contemporary Issues in Bioethics*. 3rd ed. Eds. Tom L. Beauchamp & LeRoy Walters. Belmont, California: Wadsworth Publishing Company, 1989.

Waldron, Jeremy. *The Right to Private Property*. New York: Oxford UP, 1988.

Comments on *Donald Davidson: Meaning, Truth, Language, and Reality*[1]

Simon Evnine, *University of Miami*

Donald Davidson is, notoriously, a philosopher who has appeared as many things to many people. He has been interpreted in the light of Husserl, Heidegger, Derrida and the American Pragmatists. The Davidson who appears in the pages of Lepore and Ludwig's book is none of these exotic things. They offer a straight-ahead, analytic Davidson and their rendition is, I think, pitch perfect. Not only do they get Davidson right in general terms, but on every disputed point of exegesis, their careful reconstructions of Davidson's views are exactly right. This book is, therefore, a very impressive, and a very useful achievement. It will provide the essential reference point for all future discussions of Davidson's work on, as their subtitle says, meaning, truth, language and reality. (They touch only lightly on action theory and not at all on anomalous monism or the nature of events.)

The authors' account of Davidson's *oeuvre* falls into three parts which, roughly speaking, follow the chronology of Davidson's papers, thus giving us an early, middle and late Davidson. The early Davidson, of the late 1960s, set the course for everything that was to come. He was concerned with the nature of meaning, but, as he says, he "became frustrated with the fatuity of the attempts at answers I found in Ogden and Richards, Charles Morris, Skinner and others. So I substituted [for the question "what is meaning?"] another question which I thought might be less intractable: What would it suffice an interpreter to know in order to understand the speaker of an alien language, and how could he come to know it?" Davidson's answer to the first part of his question, what would it suffice an interpreter to know in order to understand the speaker of an alien language, was a Tarski-style truth theory for the language.

Lepore and Ludwig actually discern two distinct issues Davidson was getting at with his invocation of Tarski. What they call the 'original project' was simply to show how one could specify the meanings of a potential infinity of sentences on the basis of something finite. Lepore and Ludwig suggest that if one takes the axioms of a truth-theory, axioms which say things like "'Caesar' refers to Caesar" and "'is ambitious' is true of something iff it is ambitious", as being interpretive – that is, as in some sense saying what the object language terms mean – then the truth theory shows how one can generate interpretive theorems for any object language sentence.

This is not, of course, how Davidson ever put things. The reason is that the proposal for how to specify the meanings of the infinite number of sentences in a language depends on treating

the axioms of the theory as interpretive. This is to rely on some notion of meaningfulness which is taken as primitive. The use of the truth-theory shows merely how to go from the finite base to the meanings of an infinite number of sentences. It does not show what it is for theorems or axioms to be interpretive. The attempt to use a truth-theory to accomplish this latter task is what the authors call the extended project and it is this that occupies most of Davidson's attention in the philosophy of language. (The original project returns in a companion to this volume, *Donald Davidson's Truth-Theoretic Semantics*.)

How can a truth theory fulfill the extended project? The authors show the evolution of Davidson's thinking on this. At first, he thought that if one confirmed enough theorems, of the form "'s' is true iff p," that would be enough to treat those theorems as interpretive – as saying that "'s' means that p." But it quickly became apparent, to Davidson and his critics, that this was much too weak. Confirming enough such theorems, and projecting a set of axioms from which those theorems and others could be generated, would indeed yield a well-confirmed truth theory for the object language, a theory that correctly determines the extension of the predicate "is true" over sentences of the object language. It would not show that the theorems of the theory could be taken as saying anything about meaning.

Searching for what one would have to know of a truth theory, besides its extensional adequacy, if it were to be taken as interpretive, brought Davidson to his middle period, roughly, the 1970s, the era of radical interpretation. Davidson's commitment to the idea that meaning is essentially a public phenomenon, what the authors call the primacy of the third person point of view, led Davidson in the direction of trying to show how publicly available evidence, behavioral evidence, could be used to confirm a truth theory. Confirming a theory on the basis of such evidence alone is the project of radical interpretation. But how exactly is this supposed to work? The authors reveal a certain confusion in Davidson's views on this. The question is whether confirmability by the radical interpreter is a restriction on how a truth theory must be confirmed if it is to be interpretive, or whether what must be confirmed is that the theory can be confirmed by a radical interpreter. Once this question is put, it seems clear that neither answer can be correct. If, from the point of view of a radical interpreter, one confirms that a given truth theory is indeed correct for a given language, then all one knows is that that truth theory is true, i.e. extensionally adequate. But as we indicated above, mere extensional adequacy is insufficient to guarantee that a truth theory so confirmed is interpretive. On the other hand, why should a radical interpreter's (or anyone's) showing that a truth theory can be confirmed by a radical interpreter add anything to showing that the truth theory is extensionally adequate that is relevant to its being interpretive? Ludwig and Lepore's response is that what Davidson should have said is that a radical interpreter

must be able to confirm a truth theory and confirm that it is interpretive, thus picking up on the idea they introduced in redescribing Davidson's original project of providing a compositional theory of meaning.

Presumably, at least one reason that Davidson did not put matters thus is that it appears to beg the question. If a radical interpreter can confirm that a truth theory is not just extensionally adequate but also interpretive, is she not already employing a semantic concept of the kind that is supposed to be the outcome of radical interpretation, rather than its enabling condition? Ludwig and Lepore argue that their suggestion does not beg the question, though. Semantic concepts may not enter into the description of the evidence the radical interpreter uses to confirm the truth theory (and confirm that it is interpretive); they may, however, enter into a statement of what the radical interpreter is trying to confirm on the basis of that evidence. I do wonder, though, what is involved in showing that a truth theory is interpretive. The authors seem to reject the view that showing that the truth theory can be confirmed from the point of view of a radical interpreter suffices to show that it is interpretive. I am not sure, however, what criterion of adequacy they substitute.

The method of radical interpretation is described in fascinating detail. Exactly what form the evidence takes and how one gets from that evidence to a truth theory, via the principle of charity are carefully discussed and laid out perspicuously. Ultimately, however, the authors argue that Davidson fails in his attempt to show that a radical interpreter could confirm that a given truth theory was interpretive of a speaker. Thus, the extended project attributed to the early Davidson must also fail. Davidson does not succeed in showing what it is for language to be meaningful.

The third part of the book, dealing with the late Davidson (from about 1980 onwards) deals with consequences that Davidson thought followed from his views about radical interpretation, and also with some arguments that might be thought to rebut the authors' conclusion that radical interpretation is impossible. The consequences include rejection of alternative conceptual schemes, an account of a person's knowledge of her own mental states, a rejection of skepticism and the necessity of language for thought. All of these famous positions associated with Davidson are also judged to fail, since they all, in one way or another, presuppose the possibility of radical interpretation. Davidson's well-known triangulation argument, which might be hoped to take care of the authors' arguments against the possibility of radical interpretation, also does not succeed.

The heart of the book, then, is the authors' argument that radical interpretation is not possible. Their reason for thinking this is that the whole idea of using radical interpretation, in which a person tries to arrive at a theory of what another means and believes on the basis only of publicly available (i.e. behavioral) evidence, depends on seeing the theoretical concepts of an interpretation theory (the concepts of meaning, belief, etc.) as what the authors call "purely theoretical." That is,

their significance must be taken to be exhausted by what they imply for the domain of behavioral evidence that determines when they are applied or not. Advocacy of the pure theoreticalness of the concepts of meaning and belief counts as commitment to the adequacy of the third person point of view. The reason, then, why the authors find the possibility of radical interpretation problematic is that they think that the third person point of view is not adequate. In particular, the fact that people can know about their own meanings and beliefs without the kind of behavioral evidence which is supposed to exhaust the content of these concepts from the third person point of view, shows that there is more to them than their implications for behavior. Davidson struggled to give some account of a person's knowledge of her own mental states and meanings that was consistent with his overall third-person approach. Few have found Davidson's attempted solutions to this problem at all convincing. What is novel in Ludwig and Lepore's critical treatment of this part of Davidson's philosophy is the way in which they connect Davidson' failure here to a failure in his whole philosophical project. The theory of radical interpretation is seen as part of a grand rejection of a Cartesian tradition in philosophy. But this rejection stumbles over a modest Cartesianism that just won't go away: people do know their own meanings and beliefs without evidence and with greater surety than that with which they can know the meanings and beliefs of others. Unless one can finesse that fact away, no account of meaning and belief which attempts to exhaust their significance in behavioral evidence, as radical interpretation does, can possibly succeed.

The asymmetry of first- and third-person attributions of meaning and belief has proved a thorn in the side of every theory that has ventured into that territory. I cannot help feeling that Davidson was on the right lines here, at least to this minimal extent: a solution should come from some kind of deflationism about the first-person point of view. I think that my instinct on this puts me at odds with the authors who, contrary to prevailing philosophical currents, seem to hint at a sneaking admiration for Cartesianism. About how deflationism about the first person should go, and whether it will ultimately be compatible with other Davidsonian commitments, I have no idea.

The general reason the authors have for thinking that radical interpretation must fail manifests itself at a greater level of detail in their treatment of indeterminacy and its relation to the mechanics of radical interpreation. Lepore and Ludwig argue that the evidence available in radical interpretation underdetermines an interpretive truth theory. That is, working from the admissible evidence alone, one is able to confirm a wide range of distinct truth theories equally well. Davidson, of course, has expressed allegiance to the notion of indeterminacy of meaning. It might, therefore, be thought that the underdetermination that the authors detect is simply an instance of indeterminacy and therefore not harmful to the Davidsonian project. Ludwig and Lepore's case against the possibility of radical interpretation depends on characterizing indeterminacy and

distinguishing it from the underdetermination of truth theory by evidence which they find. Indeterminacy is a special case of underdetermination. It occurs when all the relevant evidence fails to favor one of a number of theories and the concepts in those theories are purely theoretical, in the sense we specified above. In that case, the authors say, the several theories will "express the same facts." In *mere* underdetermination the various theories will not express the same facts.

Since, as we have just seen, the authors disagree with Davidson's view that the concepts of an interpretation theory, the concepts of belief and meaning, are purely theoretical, it is clear that they will take any underdetermination allowed by the method of radical interpretation to be of the non-indeterminacy variety. Supposing that radical interpretation does allow mere under-determination, its use will lead to a number of different theories which do not "express the same facts." Each of these theories will have the form of a truth theory for the language under investigation. If someone knows that the totality of evidence is compatible with distinct truth theories for the language then, even if it is permissible arbitrarily to adopt one of these theories as a *truth* theory for the language, it could not be taken as an *interpretive* truth theory. Hence, if mere underdetermination is possible in radical interpretation, it cannot succeed.

How is underdetermination possible in radical interpretation and why must it be of the non-indeterminacy variety? Here is a crude reproduction of the authors' long and complex argument. The radical interpreter starts with behavioral evidence and on that basis attempts to confirm a set of L-sentences for a speaker. L-sentences say, roughly, under what conditions the object of interpretation holds true a given (occasion) sentence. When we have a set of L-sentences, we use the principle of charity to infer from them a set of TF-sentences, which say under what conditions those occasion sentences *are* true. From the TF-sentences, we project a set of axioms that articulates their compositional structure, and the set of TF-sentences plus those axioms constitute a truth theory for the language we are interpreting. We can then take this truth theory as interpretive, if we know how it was arrived at (and perhaps some other things, such as that its concepts are purely theoretical – I'm not sure about this point, as I mentioned above).

Undertermination may creep in at various points. For example, when projecting the axioms of a truth theory from the confirmed TF-sentences, there may be several ways to do that. So long as the concepts of reference and satisfaction are taken as purely theoretical, however, such underdetermination will be harmless indeterminacy. The problem comes, according to Ludwig and Lepore, over the possibility of many, non-equivalent sets of L-sentences an interpreter can infer from the behavioral evidence. (They show how if there is one set, there will be many.) There will thus be non-equivalent TF-sentences and non-equivalent truth theories. Our knowledge of the

existence of (or the possibility of the existence of) such alternative truth theories in turn will prevent us from taking any of them as interpretive.

Why cannot this be treated as a case of benign indeterminacy? Ludwig and Lepore write:

> It is not clear from the point of view of the radical interpreter that *he* can treat different theories he could confirm as stating the same fact, which is required in order to treat the resulting range of theories as an expression of… indeterminacy… rather than simply underdetermination [of the non-indeterminacy variety]… From the point of view of the interpreter, the different L-sentences he can confirm are non-equivalent…[This will] generate corresponding TF-sentences which from the interpreter's perspective must be treated as assigning different interpretations to the speaker's sentences. Thus, the interpreter must regard the different theories he can confirm as strictly incompatible with one another (not just apparently or intuitively)… In other words, it is incoherent for the interpreter to regard the different theories which he could confirm as both true. (239)

Their point is that the *interpreter* knows that different starting points (different sets of L-sentences) will lead to different TF-sentences. These TF-sentences are what will be used to interpret the sentences of the interpretee. But they use, on their right-hand sides, sentences of the interpreter's language which the interpreter herself knows to mean different things. How, then, can the interpreter take the different interpretation theories to express the same facts? The argument, therefore, hinges on the interpreter being able to know something about her own meanings that she cannot find in others on the basis of radical interpretation.

There is, actually, one way in which an interpreter could choose indifferently among theories she knew to be non-equivalent in interpreting something: if she took her own conceptual repertoire (or language) to be richer than the interpretee's. This is what happens in the application of intentional concepts to non-human animals. Although I know that "my feeder is home" and "S.E. is home" mean different things, the difference is irrelevant when interpreting my cat. But in interpreting other humans, this cannot work. The authors point out that the interpreter may be faced with others who speak the same language she does. In that case, the interpreter cannot take her own language to be richer than her interpretee's. Another reason, not given by Lepore and Ludwig, but consonant with their outlook, is that an interpreter must, on Davidson's view, take the objects of intepretation to be able to interpret back. If a condition of interpretation were that the interpreter must take her own language to be richer than the interpretee's, this would lead to a

conundrum: A's taking distinctions in her language to be irrelevant in interpreting B, while also holding that B is taking distinctions in *her* language to be irrelevant in interpreting A.

As I stated above, the authors take themselves to be showing that radical interpretation entails a kind of underdetermination that is not of the benign indeterminacy variety. But Davidson describes indeterminacy as allowing for "equally acceptable alternative theories which differ in assigning clearly non-synonymous sentences of [the interpreter] as translations of [the interpretee's] same utterance." The authors, therefore, might better have put their point by saying that, contrary to Davidson's claims, indeterminacy is not benign after all. And the reason it is not benign has to do with the fact it will represent as irrelevant differences that an interpreter will know are relevant, in application to herself.

The book, after 420 dense, difficult and fascinating pages, ends abruptly, in this sense. Having argued that the project of radical interpretation must fail, and with it a variety of arguments Davidson pressed on its basis, the authors do not take any final stock of Davidson's contribution to philosophy. If the authors are right, almost everything Davidson wrote on meaning, truth, language and reality is flawed. (The same is true, of course, of all interesting philosophers.) What, then, should we make of Davidson? What are the lessons to be learned? How does Davidson's project fit into the bigger picture of 20th century philosophy? It is, perhaps, understandable that, having stared so hard and so perceptively for so long at the details of Davidson's *oeuvre*, the authors should have found it difficult to refocus their gaze on such Brobdingnagian questions. But Lepore and Ludwig's already excellent book might have been even better had they done so.

Note

[1] Reprinted with kind permission from Oxford University Press, from *Mind* 116 (April 2007): 446-53.

Comments on *Donald Davidson: Meaning, Truth, Language, and Reality*

Piers Rawling, *The Florida State University*

What follows is a version of comments presented at the November 10, 2006 meeting of the Florida Philosophical Association. I have modified my original remarks somewhat in light of input at the meeting, and I would like to thank all those who participated – particularly Kirk Ludwig. I owe a far larger debt to Donald Davidson himself, who was my dissertation supervisor, and spent many hours explaining his views to me, and attempting to teach me to think philosophically. He is, of course, not responsible for any errors herein.

I shall endeavour to make my remarks accessible to those unfamiliar with his work. In doing so, I will no doubt sacrifice subtlety and accuracy for simplicity. But if you want the full dose of the former pair, I suggest you consult *Donald Davidson: Meaning, Truth, Language, and Reality*. This is an excellent book that covers and critiques just about everything Davidson ever wrote. And there is much in it with which I agree. So much, in fact, that I am not going to venture to disagree with anything Ernie Lepore and Kirk Ludwig (hereafter EK) say. Rather, I am simply going to bolster one of their lines of argument with one of my own. In doing so, I'll be drawing on Rawling 2001 & 2003; the former is followed by a reply from Davidson[1]; the latter is in Kirk Ludwig's edited volume *Donald Davidson* – so much of this will be familiar to him, but perhaps others may benefit.

The argument concerns whether or not radical interpretation is possible. EK argue that it is not. I agree. But I have a somewhat different argument for its impossibility.

Radical Interpretation

First, what is radical interpretation? Suppose you find yourself stranded on a desert island with another castaway, and rapidly discover that you cannot communicate with her; indeed, it is perhaps not clear initially that she has a language. You must build your interpretation of her, as agent and speaker, from scratch. It is interpretations such as this that are radical.

Davidson, following, to a certain extent, Quine, is famous for pursuing the issue of how such radical interpretation is possible. How is this pursuit undertaken? By exploring the constraints on interpreting the raw data of overt behaviour – constraints that we must respect if our enterprise is to be the interpretation of some other being as a speaker and a thinker. Thus, in understanding these constraints, we can come to learn much about what is required to be a speaker and a thinker.

These constraints involve assumptions that the interpreter must make about the interpretee. Davidson argues, then, that for interaction to be mutual interpretation, the parties must make assumptions about each other that could not turn out to be false lest their enterprise fail to be interpretation at all. In a sense, then, no interpretation is built entirely from scratch, and it is this that makes radical interpretation (hereafter RI) possible.

To avoid confusion, it should be noted that Davidson's exercise is conceptual. What emerges is not a manual for the field linguist, but a distinctive, and in many respects compelling, picture of language and the mind. Importantly, Davidson sees as impossible the conceptual reduction of our intentional concepts (those we use to describe the meanings and minds of others) to the non-intentional. He seeks, rather, to illuminate the former by examining their relations to one another and to the evidence we use in their deployment.

Charity

What are the constraints on RI – the constraints that make it possible? These are summed up under the 'Principle of Charity'. This is a term borrowed by Quine from N. L. Wilson[2]; and it is a purposeful misnomer. There is nothing charitable about its application: it must be applied if we are to be engaged in interpretation.

There is controversy, however, over its best formulation – EK have a very illuminating discussion of it (p. 185 ff) But the basic idea is that the interpretee must be interpreted (1) as being largely rational, and (2) as having largely true beliefs (about her immediate macroscopic environment, at least). For instance, in the absence of evidence that your interpretee is round the bend, an interpretation that has him making inexplicably crazy remarks is likely to be *mis*interpretation. Much more needs to be said to fill out this principle, of course. Which kind of beliefs, and what proportion (can this really be quantified?) of them, must be true? And what exactly is rationality? Does it go beyond, say, a certain amount of logical acumen and some basic constraints on one's preferences? I shall not address such issues here: the interested reader can consult EK for more detail.

Indeterminacy

One of Davidson's central claims is that the constraints on interpretation are not sufficient to force uniqueness: interpretation is inevitably indeterminate (see EK chapter 15 and the Davidson references therein). There will be many interpretations (perhaps infinitely many) that respect the

principle of charity and account for the behavioural data (non-intentionally described, of course – that is, the behaviour is not to be described using propositional attitude vocabulary since that already is to presume an interpretation). As Quine puts it for translation: "translation [is] indeterminate in principle relative to the totality of speech dispositions."[3] And Davidson sees the same holding true for interpretation (the difference between translation and interpretation is, roughly, that to interpret someone is to understand them, whereas you might know that some sentence is, say, a French translation of some German sentence without understanding either). Radical interpretations, then – including attributions of beliefs – are not uniquely determined by the evidence.

The basic idea of this sort of indeterminacy is familiar from discussions in the philosophy of science: the data, it is said, underdetermine theory. Many different theories can account for the observations made in any particular realm – and this is true, perhaps, even in the ideal limit when we have made all the observations that are humanly possible. One sort of instrumentalist claims, in light of this, that we should abandon the quest for truth when it comes to scientific theories that go beyond our observational capacities. Rather, such theories should be viewed as convenient instruments that aid us in making accurate observational predictions about, and successful interventions in, our world. Such a view may be tied to an epistemic conception of truth, according to which the truth cannot outrun our ability to know it. If observation is required for scientific knowledge, and theory outruns our powers of observation, then theories cannot be known. If in addition truth-value is tied to knowledge, then theories lack a truth-value – it is not merely that their truth-values cannot be known; rather, they have no truth-value. We might say that their truth-values are not merely epistemically indeterminate, but metaphysically so.

Indeterminacy, then, enters this account of instrumentalism in two places. First there is the claim that observation does not determine theory. And second it is claimed that truth-value is metaphysically indeterminate in the absence of knowledge. Compare this with Davidson's account of the indeterminacy of interpretation. So far we have a parallel to the first indeterminacy: there are many interpretations that account for the behavioural data. But Davidson goes further to give us, apparently, a parallel to the second indeterminacy – for example "meaning is entirely determined by observable behaviour."[4] And the same goes for all the propositional attitudes – there's 'no more' to, say, beliefs than what can be gleaned from radical interpretation, and radical interpretations are not unique.

So now it looks as though Davidson must abandon claims to the effect that we really do have propositional attitudes – just as the instrumentalist above abandoned truth-values when it comes to theory, so Davidson must abandon truth-values when it comes to, say, remarks to the effect that so-and-so believes that *p*. Or, to put it more starkly, there are no beliefs, nor any other

propositional attitudes for that matter. The indeterminacy of interpretation is not only epistemic, but also metaphysical.

But now we have:

> (1) Radical interpretation requires charity (this is what makes RI possible by ensuring that radical interpretations are not constructed entirely from non-intentionally described behaviour alone). And charity requires the existence of beliefs: not only must the interpretee have beliefs, but many of them must be true.
> (2) There's 'no more' to beliefs than what can be gleaned from radical interpretation.
> (3) But radical interpretations – including attributions of beliefs – are not uniquely determined by the evidence.

Together, (2) & (3) entail that there are no beliefs. (1) entails that radical interpretation requires beliefs.

If, then, to say that radical interpretation is possible is in part to say that we have beliefs and they can be determined from the stance of the radical interpreter, then radical interpretation is impossible. Davidson, of course, would disagree. Rather, he claims that the multiplicity of radical interpretations is no more problematic than, say, the multiplicity of scales of length or temperature measurement – radical interpretation is analogous to measurement.

Does the Measurement Analogy Save the Day?

The basic idea of Davidson's analogical strategy is clear enough.[5] We use sentences to track the propositional attitudes of an agent; we use the real numbers to track, say, the lengths of objects. The sentences in the former attribution play the role of the real numbers in the latter. We might maintain that there is an 'indeterminacy of length': there are infinitely many serviceable schemes for attributing lengths (feet, inches, metres, etc., and we could come up with infinitely many more). Similarly, there are many serviceable schemes for attributing meanings and the other attitudes to a given interpretee. And the latter indeterminacy, Davidson claims, is as benign as the former. But the details of the analogy need spelling out.

Davidson does not deny that there are many disanalogies between the two cases. There is, for example, an algorithm for moving from one scale of length measurement to another (multiply by the relevant positive constant – for example, multiply by three to move from feet to yards); there is no such algorithm in the case of interpretations. However, we do need at least something invariant

across interpretations if there is to be any appropriate analogy at all. What does invariance amount to in the measurement theoretic case? First, there are invariant relata. In the case of length, for instance, we attach the numbers to physical objects (the relata), and these remain the objects to be measured as we move from one scale to the next. And each scale preserves the relation 'is at least as long as': the greater the number on any given scale, the longer the object; and if it's longer on one (say feet), it's longer on all the others (metres, inches, etc.).

What remains invariant across radical interpretations? We needn't insist on pushing the analogy all the way, but at least we should be able to find a class of relata whose membership is fixed across different interpretations. It might initially appear that these relata are the propositional attitudes of the interpretee. But do these remain invariant across different interpretations? No. It is precisely they that vary. On Davidson's picture, indeterminacy is simply the fact that we can use different locutions to locate the same node in some pattern. But what are the invariant nodes? They cannot be propositional attitudes: the belief that *p*, say, under one scheme, will be the belief that *q* under another – two different propositional attitudes (both attributions are couched in the one idiolect of the interpreter). The relata cannot comprise the propositional attitudes themselves since we identify propositional attitudes in part by their content. And their content varies between schemes of interpretation.

One might respond here by wondering whether there is not some 'neutral' way of identifying propositional attitudes – so that, despite appearances to the contrary, the belief that *p* and the belief that *q* are in fact the same belief. But the problem is worse than so far suggested. Propositional attitudes in one scheme need not even map one to one onto propositional attitudes in another. Consider the following possibility: on one interpretation, a piece of behaviour, *B*, is interpreted as a signal; on another interpretation (that also saves all the relevant phenomena), *B* is interpreted as a simple scratch.[6] Thus, on the first interpretation, *B* is explained by a complex of propositional attitude states that is simply larger than that invoked in the second. Or, indeed, the explanation of *B* on some scheme might invoke no propositional attitudes at all.

What is Invariant?

What, then, are the invariants across interpretations on Davidson's account? Quine has us tracking "men's dispositions to respond overtly to socially observable stimulations."[7] And according to Davidson, "… a satisfactory theory is one that yields an acceptable explanation of verbal behaviour and dispositions."[8] Behavioural dispositions are key; and they are apparently the invariant nodes across different schemes of interpretation – they are what propositional attitude attributions

'measure'. Just as we attribute numbers in the form of lengths to track the relation 'is at least as long as', so we attribute propositional attitudes to track behavioural dispositions.

The invariants, then, have the following form: a disposition to *B* under circumstance *C*, where *B* is a bodily movement non-intentionally described, and *C* is specified without reference to propositional attitudes of the agent. A full interpretation of an agent – something unattainable by mere mortals – would account for the totality of the agent's behavioural dispositions. And all adequate interpretations must agree on these dispositions; but (modulo charity, of course) that is all they must agree on. Propositional attitude states seem to be mere posits; they vary from interpretation to interpretation. That we cannot re-identify a belief across interpretative schemes is a consequence of the fact that there is no such state to re-identify.

Davidson's Instrumentalism?

To return to instrumentalism, an instrumentalist in the philosophy of science holds something akin to the following. The constraints on a theory are that it should save the phenomena and meet some criteria of coherence and simplicity (and elegance, perhaps). Theories will (usually) posit entities that underlie the appearances; but, according to the instrumentalist, these are only posits – constructions designed to save the phenomena. If two theories were to save the phenomena, score equally well in other respects, but disagree as to posits, the instrumentalist would have no difficulty in countenancing both: the posits, being merely that, would not compete.

Davidson's view of the mind, then, might appear to be instrumentalist, the relevant criteria of coherence being provided by the principle of charity; the posits comprising, amongst other things, propositional attitudes; and (non-intentionally described) behaviours and behavioural dispositions constituting the phenomena to be saved. About some posits Davidson is quite explicit:

> words, meanings of words, reference, and satisfaction are posits we need to implement a theory of truth. They serve this purpose without needing independent confirmation or empirical basis …. [Satisfaction and reference are] notions we must treat as theoretical constructs whose function is exhausted in stating the truth conditions for sentences.[9]

And:

> degree of belief is a construction based on more elementary attitudes.[10]

When it comes to the propositional attitudes he seems, however, more wary. He does say:

> In thinking and talking of the weights of physical objects we do not need to suppose there are such things as weights for objects to have. Similarly in thinking and talking about the beliefs of people we needn't suppose there are such entities as beliefs.[11]

But in the same paper, he speaks of beliefs as mental states with identity conditions.

However if Davidson is correct in his views that

> (i) interpretation is epistemically indeterminate: many different interpretations will save the phenomena; and
> (ii) all of these will be correct, since there is no more to correctness here than saving the phenomena

then there are no propositional attitudes. (i) implies that the propositional attitudes are epistemically underdetermined in the sense of out-running the evidence. Yet (ii) implies that there is no more to the propositional attitudes than the evidence provides, so that the evidence cannot be out-run. But nothing can fit both these bills.

Can Davidson, however, really adopt instrumentalism concerning the propositional attitudes? Is the position even coherent? The first difficulty is that the principle of charity would have to be reformulated – it could no longer commit us to the existence of beliefs, let alone true ones. And then there is a more overarching concern that there would be no perspective from which the view itself could be stated since there would be no statements. Furthermore, we couldn't know any of this since (as Eve Garrard pointed out to me) there would be no states of knowledge.

There are also more specific problems for others of Davidson's own views. For example, he relies upon the existence of propositional attitude states in his (justly famous) causal theory of action – roughly speaking, they're what does the causing. I'll conclude with a brief discussion of another aspect of his views, however.

Davidson's Anti-Cartesianism

There is an anti-Cartesian thread that runs through Davidson's work on interpretation.[12] As EK put it:

> If Davidson is right, then the central mistake of the philosophical tradition is the assumption of the Cartesian standpoint, and, in particular, the central place in epistemology accorded to the epistemic priority of knowledge of our mental states to knowledge of the world and other minds (EK, p. 418).

The Cartesian epistemic perspective is first-person: you start from knowledge of your own mind and attempt to move outward. Davidson, by contrast, proposes (roughly speaking) to begin in the external world. When it comes to the problem of other minds, for example, one way to look at matters is to see Davidson as putting the shoe on the other foot: the problem is not so much one of other minds as a problem about your own. The minds of others are accessed through interpretation of their behaviour, but how do you access your own mind? It's not as though you observe your own behaviour and compile a self-interpretation (or, at least, such goings on are very much the exception).[13] This anti-Cartesianism is heady stuff, and potentially revolutionary. But, as EK go on to point out (pp. 418-419):

> The central difficulty, we believe, lies in the apparent underdetermination of the content of our thought about meaning and thought content by behavioural evidence, including interactions with objects in our environment. What must be done is to show that the [external] standpoint can yield acceptably determinate assignments of contents to beliefs and interpretations to sentences.

Yet again, I agree. One could hold that there really are propositional attitudes and there's nothing more to them than is accessible to the radical interpreter if it could be shown that radical interpretation is not epistemically indeterminate (or, at least, is far less so than Davidson argues in some places[14]).

Notes

[1] Donald Davidson, "Reply to Karlovy Vary Papers," in *Interpreting Davidson*, eds. Petr Kotatko, Peter Pagin & Gabriel Segal (Center for the Study of Language and Information, 2001), 298-300.

[2] See W.V.O. Quine, *Word and Object* (Cambridge, Mass.: MIT Pr, 1960), 59.

[3] Quine 221.

[4] Donald Davidson, "The Structure and Content of Truth," *Journal of Philosophy* 87 (1990): 314. This is discussed by EK on p. 166 ff.

[5] See, e.g., Donald Davidson "What is Present to the Mind?" in *Subjective, Intersubjective, Objective: Philosophical Essays Volume 3* (Oxford: Clarendon Pr, 2001).

[6] This example is similar to one suggested to me by Kirk.

[7] Quine ix.

[8] Donald Davidson, "The Inscrutability of Reference," in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Pr, 1984), 237.

[9] Donald Davidson, "Reality without Reference," in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Pr, 1984), 222-223.

[10] Davidson, "The Structure and Content of Truth," 322. See EK, p. 225 ff for further discussion.

[11] Davidson, "What is Present to the Mind?"

[12] See his "What is Present to the Mind?" for example.

[13] See, e.g., Donald Davidson, "First Person Authority," in *Subjective, Intersubjective, Objective: Philosophical Essays Volume 3* (Oxford: Clarendon Pr, 2001) and EK, chapter 20, and the references therein.

[14] See, e.g., Davidson, "The Inscrutability of Reference."

Works Cited

Davidson, Donald. *Inquiries into Truth and Interpretation.* Oxford: Clarendon Pr, 1984.

Davidson, Donald. "The Structure and Content of Truth." *Journal of Philosophy* 87 (1990): 279-328.

Davidson, Donald. *Subjective, Intersubjective, Objective: Philosophical Essays Volume 3*. Oxford: Clarendon Pr, 2001.

Davidson, Donald. "Reply to Karlovy Vary Papers." In *Interpreting Davidson*, edited by Petr Kotatko, Peter Pagin & Gabriel Segal, 285-307. Center for the Study of Language and Information, 2001.

Quine, W.V.O. *Word and Object*. Cambridge, Mass.: MIT Pr, 1960.

Rawling, Piers. "Davidson's Measurement-Theoretic Reduction of the Mind." In *Interpreting Davidson*, edited by Petr Kotatko, Peter Pagin & Gabriel Segal, 237-255. Center for the Study of Language and Information, 2001.

Rawling, Piers. "Radical Interpretation." In *Donald Davidson*, edited by Kirk Ludwig, 85-112. Cambridge, U.K.: Cambridge UP, 2003.

The Problem with Radical Interpretation: Reply to Simon Evnine and Piers Rawling

Kirk Ludwig, *University of Florida*

I want to thank Simon and Piers for their comments. It is real pleasure to have such perceptive commentators. Both Simon and Piers put the argument for the impossibility of radical interpretation at the center of their comments, and I will do the same with the aim of shedding more light on where the fundamental difficulties lie.

Radical interpretation is the centerpiece of Davidson's philosophy. It incorporates his suggestion for using a Tarski-style truth theory, adapted to a natural language, to do the work of a compositional meaning theory. It puts at center stage Davidson's conviction that meaning and related matters must be understood from the third person point of view because of the essentially public nature of language and communication. It exhibits his commitment to explicating the family of concepts involved in understanding language not by giving analyses one by one but by trying to show how evidence which does not presuppose their application can be marshaled in support of a theory deploying them. And in arguing that radical interpretation represents, in some suitable sense, the basic methodological and epistemic position of one speaker with respect to another, Davidson puts himself in a position to argue that whatever must be assumed for radical interpretation to succeed is constitutive of the interpreter's subject matter.

There is no denying the essentially public character of language. The central question about Davidson's philosophical work is whether what is undeniable is strong enough to support the view that interpretation can succeed from the extremely restricted evidential position of the radical interpreter. The basic idea is that speakers are by their nature interpretable, the function of language being communication between speakers. Davidson interprets this as requiring interpretability form the standpoint of the radical interpreter. Why? We can see why this follows if we interpret the publicity requirement as in (1).

(1) Any speaker is interpretable in principle by any other speaker in any environment.

This may seem very strong, but if this is true, then given the deployment of the principle of charity in interpretation, we get immediately that any speaker must be mostly right about his environment,

that there can be no radically different conceptual schemes, and that what is not recoverable from purely behavioral evidence is no part of meaning – all theses which Davidson has endorsed.

Lepore and I think that a weaker principle accommodates the essentially public character of language, namely,

(2) Any speaker is correctly interpretable by some speaker in some environment.

This is too weak, though, to support most of what Davidson has wanted to maintain about the essential character of language and the propositional attitudes. The stronger principle, which would underwrite the possibility of radical interpretation, we try to show runs into difficulties from the standpoint of radical interpretation itself.

The radical interpreter, if we are right about Davidson's ultimate commitments, is restricted to purely behavioral evidence. Davidson often describes the radical interpreter as starting with knowledge of a speaker's hold true attitudes. However, he also describes this as an intermediate stage of interpretation. He presupposes that the radical interpreter can identify hold true attitudes on the basis of more primitive evidence. It is important that in describing the evidence as purely behavioral, we are excluding any knowledge of the speaker's psychological states, and any general knowledge about the type of psychology he has. All of that has to be justified in terms of the behavioral evidence. And it is not in the cards, on Davidson's view, to appeal to another speaker being a conspecific and to infer from knowledge of one's own mind what the other's mind, adjusting for different histories, is likely to be like on that basis. This would be to make the first person point of view in an important sense primary in the epistemology of interpretation.

In approaching the question whether radical interpretation is possible, as Simon noted, we identify as a fundamental assumption of Davidson's project that the concepts invoked in the theory of interpretation are purely theoretical. They are purely theoretical if their import is exhausted in their role in the theory in accounting for the behavioral evidence for it. We challenge this on the basis of the observation that we seem to apply the concepts of the theory both on the basis of behavior and directly to ourselves. This suggests prima facie that their content is not thought of by us as exhausted by their role in imposing an intelligible pattern on behavior in interpretation. If Davidson had a way of making sense of this compatibly with his overall position, then he would have a response to this objection. We argue that his accounts fail. If the concepts are not purely theoretical, then a real question arises about which of two incompatible theories equally compatible with the evidence is correct. The claim of indeterminacy in interpretation, as we see it, rests centrally on Davidson view of the concepts of the theory of interpretation being purely theoretical.

However, our objection to the possibility of radical interpretation does not rest solely on this. We also aim to raise a difficulty that is more internal to the project, though, as it turns out, one that is ultimately connected in an important way with the issue just raised, as Simon brings out.

It will be helpful to begin by laying out a principle which our argument can be seen as resting on:

> (3) If there can be genuine indeterminacy of theories of interpretation, it must be possible for this to make sense from the standpoint of an interpreter of another speaker.

Why impose this requirement? First, it would be odd if we maintained that there was indeterminacy of interpretation, but when we actually engaged in interpretation, we could not accept that that was so. But, second, it should follow from Davidson's own fundamental assumption that the most basic methodological stance we can take with respect to language and related matters is the standpoint of the radical interpretation. If indeterminacy cannot emerge from that standpoint, we should, by that methodological principle, reject it.

What we argue is that there is at least one stage in the interpretation of another speaker, in accordance with the procedure which Davidson describes, where there are different starting points for projecting a truth theory for the speaker's language, which cannot be taken by the interpreter as saying the same thing about what sentences in the object language mean. The basic difficulty arises because the interpreter must confirm correlations between conditions in the speaker's environment and what sentences the speaker holds true in response. This gives rise to sentences of the form (L). At the next stage, one assumes (subject to correction in the light of more evidence and holistic considerations) that this provides a target interpretive T-sentence for a truth theory for the speaker's language of the form (T) (since the speaker holds true *s* iff he believes that *p* and *s* means that *p*).

> (L) *S* holds true *s* at *t* iff *p*
> (T) *s* is true iff *p*

But the causal structure of the world being what it is, there will be many different correlations or sets of correlations to choose from. And that means, from the point of view of interpretation, many different incompatible starting points, for from the interpreter's point of view, the different correlations are with different states of affairs, and so assigning meaning on the basis of one will be to say a sentence in the speaker's language means something different from assigning it on the basis of another. This is underdetermination, not indeterminacy, from the point of view of the

interpreter, subject to one caveat, which I come to in a moment. Invoking our principle above, we should conclude that, if these starting points yield theories that are empirically equivalent, radical interpretation is impossible—again subject to one further caveat.

What is the caveat? Well, how could we make out that despite how things look from the interpreter's standpoint with respect to the differences between the theories he deploys what we really have is indeterminacy? We would have to maintain that each of the theories kept track of the same facts equally well. We could make sense of this by appeal to Davidson's analogy with measurement theory if we held that what is going on is that the structure the interpreter uses to keep track of the facts of the matter with respect to the speaker, like the numbers in application to physical quantities they are used to keep track of, is essentially richer than the phenomena which the interpreter is keeping track of. When we use the numbers to keep track of temperatures, for example, we do not use all of their features. For the Centigrade and Fahrenheit scales, there is no sense given to an absolute origin. One can choose to assign 0 to any temperature. Then one can choose to assign 100 to any other temperature to fix the relative positions of all temperatures with respect to one another. Since sameness of differences can be physically made sense of with temperatures, we get a scale which is unique up to the choice of origin and a linear factor. There are different ways of mapping numbers onto temperatures because a subset of the features of the number is used in keeping track of the physical phenomena we are concerned with.

This would make sense of indeterminacy from the interpreter's standpoint, despite the point made above. But this move has a fatal flaw. The problem is that it requires the assumption that the interpreter's language can mark distinctions that cannot be marked in the speaker's language. But radical interpretation is supposed to be possible with respect to any language. A speaker of the interpreter's language would possess a language as rich as the interpreter. That fact could not emerge from the standpoint of radical interpretation if we employed this maneuver. And so radical interpretation fails, and the one maneuver that could rescue it fails as well because it cannot be applied to the interpreter's own language.

How is this problem related to the assumption about the theoretical character of the concepts of the theory of interpretation? Here is what we think: that assumption must be consistent with what emerges from the standpoint of the radical interpreter, on the assumption that the evidence the concepts are to keep track of is purely behavioral. That is, it can't look as if there is genuine underdetermination from the interpreter's standpoint if that assumption is true. But it does look as if there is genuine underdetermination from the interpreter's standpoint. Therefore, the assumption (relative to the evidence noted), is incorrect.

Simon draws attention to a connection between this argument and our rejection of the view that the concepts deployed in interpretation are purely theoretical concepts, in the sense we have noted. The problem we noted was that we deploy the concepts from the first person standpoint as well, and it is clear that they are not there deployed on the basis of how they keep track of patterns of behavioral evidence. The first person standpoint shows up in the argument we have just given as well. For it is the interpreter's first person standpoint on his own meanings which gives rise to the view that the speaker he is interpreting does not mark in his language distinctions as fined-grained as those that can be marked in the language of the interpreter. And so in a way the fundamental difficulty for Davidson emerges as the fact that the third person point of view presupposes a first person point of view, that of the person who takes the third person point of view toward another. When the person taking the third person point of view toward another applies concepts to himself in a way other than on the basis of adopting toward himself the third person point of view, his deployment of them to others has to respect the distinctions he marks in his first person deployment of them.

Piers offers a different, and important, argument against the possibility of radical interpretation, namely, that the Principle of Charity, which enjoins us to find the speaker's beliefs to be about what prompts them, is incompatible with the degree of indeterminacy which results from the restrictions on the evidence with which the radical interpreter must work. Piers suggests that indeterminacy with respect to the *content* of beliefs and other propositional attitudes itself entails irrealism or instrumentalism about beliefs. I think that this is not enough, because Davidson can make sense of an invariant content of a given belief across different interpretation schemes, on the model of measurement theory (though as indicated above this move has its problems too). That is, as long as he can argue that in different schemes the same belief with its content is kept track of with different sentences, each of which is acceptable because, as in the use of numbers in keeping track of physical quantities, the structure they are a part of is rich enough to be mapped in different ways on to the phenomena. However, Piers also argues that between different schemes of interpretation there is no reason to expect that it is the same states of the speaker that will be treated as interpretable, and this, I think, raises a further and deeper difficulty. If a disposition grounded in the speaker's physical constitution may under one scheme be treated as contentful but not under another, then there will not be a scheme invariant set of beliefs and other propositional attitudes for a speaker. Thus, there is no fact of the matter, for example, about whether a given objective state of a speaker is a belief or desire or whether or not it is even a propositional attitude. But as the Principle of Charity seems to presuppose that there is a fact of the matter about the speaker's

beliefs, one of the fundamental constraints on interpretation on Davidson's view seems to be in conflict with the extent of the indeterminacy the evidence and totality of constraints admit.

I agree with Piers about the seriousness of this problem. One response might be that there are enough constraints given the evidence to determine an invariant set of states as the set of a speaker's propositional attitudes, but I do not myself see how to make good on this. Supposing that the constraints and data available to the radical interpreter fail, as Piers argues, to determine a unique set of states as the set of the agent's propositional attitudes, it looks as if Davidson must embrace an instrumentalist view of propositional attitude talk which sits uneasily not just with the Principle of Charity but with much else that he has to say. Yet let me make one suggestion designed to relieve a bit of the pressure on Davidson's position by suggesting a way of construing the use of the Principle of Charity which makes it consistent with this result, even if it does not make the result consistent with everything else that Davidson says. The idea is to treat the Principle of Charity as a holistic requirement in the following sense. We aim to keep track of behavioral dispositions. There are different theories that will do this equally well, which even differ on which dispositions are treated as attitudes, and so there are no constant propositional attitude nodes across different acceptable theories. Nonetheless, we impose on each acceptable theory the following holistic constraint: it makes most of the speaker's general beliefs and beliefs about his environment true. So the idea is that we impose the Principle of Charity as a holist constraint on each acceptable theory, without presupposing identity of beliefs across different interpretation theories. Understanding its application in this way, the Principle of Charity is not in conflict with the consequence of the indeterminacy arising from the constraints on the interpreter's project.

Let me finally turn to the questions that Simon raises at the end of his remarks.

> If the authors are right, almost everything Davidson wrote on meaning, truth, language and reality is flawed. (The same is true, of course, of all interesting philosophers.) What, then, should we make of Davidson? What are the lessons to be learned? How does Davidson's project fit into the bigger picture of 20th century philosophy?

These are large questions, and I can only say some modest things here. But the first thing to be said is that there is much to be learned from Davidson's work even if we are right that the project fails to meet its ambitions. For in exploring in detail why an ambitious transformative project like Davidson's ultimately fails we inevitably gain a much deeper understanding of the domain which it aims at illuminating. Think of Kant's project in the *Critique of Pure Reason*. One does not have to suppose it to be successful to recognize how much it has deepened our understanding of reason and

the mind-world relation. Second, we can say some thing about what can be pursued independently of full success in the project. I think that Davidson is right that the standpoint of interpreting another speaker is methodologically fundamental to meaning and related matters. If we are right, he was wrong to think that it could be carried out from the very special position of the radical interpreter. But this does not mean that we can free the project of understanding language from the third person standpoint. An important follow up project is to ask what additional constraints must be supplied in order to show how interpretation can succeed on the basis of public evidence, and perhaps some additional assumptions – fundamentally, I think, that the speakers we set out to interpret are like us in certain psychological respects which helps to narrow down the range of acceptable theories of interpretation. It must still be the case that the essentially public character of language constrains understanding of meaning and language and the propositional attitudes, and much of what Davidson says about this must still be true. What we want to do now is to rethink how that fact shapes the concepts we deploy in interpretation while we free ourselves from the overly stringent requirements of radical interpretation. In addition, we regard Davidson's work in the theory of meaning, and in particular his exploitation of a truth theory in pursuit of a meaning theory, as of fundamental importance. The lessons of this have yet to be fully grasped, we think.

Finally, what is Davidson's place in 20th century philosophy? Though Davidson rejected traditional empiricism and its identification of sensory experience as the source of meaning and the epistemological foundation for knowledge, his philosophy is nonetheless a development out of the broadly empiricist movement in 20th century analytic philosophy, and to understand its historical significance it is important to understand how it grows out of this background. In that tradition which stretches back through Quine to Carnap and the Logical Positivists, and thence to the early Wittgenstein, Russell and Moore, Quine marks an important turning point. It is in terms of Quine's rejection of certain key doctrines of the Logical Positivists, and further changes Davidson made on the foundation those provided, that we can see the steps by which, through a series of changes internal to the tradition, Davidson turned the traditional picture upside down, giving primacy to a fully objective, even quotidian, the third person point view, rather than, as in the empiricist tradition, the first person point of view.

Carnap was the greatest influence on Quine, just as Quine was on Davidson. We see Quine's work most clearly in seeing it grounded in Carnap's work, though developing from in part by way of, the rejection of certain key theses in Carnap. Similarly, we can see Davidson's work most clearly by seeing it as grounded in Quine's work, though developing in part from the rejection of certain characteristic features of Quine's view. Carnap saw sensory experience as the source of meaning and as the epistemic foundation of our theories of the world, and he accepted a sharp

distinction between analytic and synthetic truths. Quine rejected the analytic/synthetic distinction of traditional philosophy as insufficiently clear, and, in rejecting it, he rejected as well the allied distinction between the a priori and the a posteriori, a distinction which Carnap had aligned with that between the analytic and synthetic. With this Quine gave up the view that philosophy was fundamentally distinct from the sciences and could provide an a priori foundation for them. He saw philosophy rather as continuous methodologically with the various disciplinary sciences and distinct largely in working with broader categories. In this case, philosophical theories become subject to the constraints which govern the development of scientific theories generally.

In rejecting the analytic/synthetic distinction, Quine likewise rejected as insufficiently clear the traditional empiricist notion of meaning. When he set out to replace the traditional notion with something more scientifically respectable in *Word and Object*, his fundamental starting point was the observation that "language is a social art." It is this that motivates the stance of the radical translator as fundamental in reconstructing a scientifically respectable notion of meaning, or at least sameness of meaning, i.e., translation. For it follows from the essentially social character of language, Quine argued, that evidence for its acquisition and deployment must be intersubjective, and, hence, recoverable from overtly observable behaviour. The concepts involved must have their content exhausted by their role in keeping track of the evidence, and, hence, two theories which accommodate the evidence equally well must have the same content. This is the application of the traditional empiricist idea of content as exhausted by contributions to confirmation conditions, at the level of the theory. In a conservative revision of the traditional empiricist theory of meaning, which keyed content to sensory experience, Quine then keyed sameness of meaning to sameness of response to patterns of physical stimulus of the sensory surfaces. This extrudes the traditional subjective basis of content in the empiricist tradition, sensory experience, to the sensory surfaces and renders it in principle intersubjectively available.

Davidson took over from Quine the idea that the third person standpoint – the standpoint of an interpreter of another, in Davidson's development of the idea – is methodologically fundamental in understanding meaning. He also took over from Quine the idea that, as language is intersubjective, the evidence to which the notion of meaning is responsible must ultimately be purely behavioural evidence, described neutrally with respect to what it may mean or what attitudes it may be connected with. Unlike Quine, however, Davidson did not set out to introduce a new more scientifically respectable notion of meaning, but to understanding and describe the actual workings of language. Meaning as it is ordinarily understood, he held, was a notion recoverable from ordinary behavioural evidence, neutrally described. This is connected with two important, connected, differences between Quine's approach to the theory of meaning and Davidson's. The first is that

Davidson treats a theory of interpretation as including a full theory of a speaker as a rational agent whose propositional attitudes organize its behaviour into intelligible patterns. The second is that Davidson takes one step further the transformation of the basis of shared meaning in Quine from the subjective to the objective by pushing the shared basis of meaning out from the sensory surfaces to distal events and objects in the environment. In so doing, he rejects the last vestige of empiricism in Quine's philosophy.

The upshot is that the basis of meaning, thought and knowledge in Davidson's philosophy is the intersubjectively shared distal environment. Knowledge of our own minds, the minds of others, and the external world are part of the basic package that makes thought and meaning possible. This vision of our embedding in the world is as far from traditional empiricism as could be imagined. It is, in a certain direction, as this tracing of the stages of its development shows, the culmination of a line of response initiated by Quine to empiricism as represented by the Logical Positivists, and centrally by Carnap. In a very loose analogy, we may think of the line of development Carnap-Quine-Davidson as similar to that of Locke-Berkeley-Hume in the modern period, in the sense that this development in the history of ideas, in retrospect at least, appears to be almost determined by a logic internal to the philosophical problems they were responding to and their particular position in the historical development of ideas about them.

I end with a word about the structure of the problem area, the mind-world relation, to which both traditional empiricism and Davidson's philosophy is a response. Each can be seen as taking a different starting point and thereby rendering unproblematic for itself something that emerges as problematic for the other. The most difficult challenge for empiricism is the recovery from the subjective foundation it provides an objective world as an object of knowledge. The corresponding difficulty for the intersubjective foundation Davidson offers is how to make sense of knowledge of the subjective, in the sense of knowledge of our own thoughts and meanings from the first person point of view. It is not a challenge that Davidson overlooked or failed to respond to. But part of the burden of our book is that the challenge has not been adequately met. And then the larger question is whether there is any position which can accommodate the role that the first person plays in our self-conception compatibly with making sense to ourselves of our having secured reflective knowledge of a fully objective mind-independent world.

Notes on Contributors

Nathan Andersen is an Associate Professor of Philosophy at Eckerd College, in Saint Petersburg, FL. He focuses on the history of philosophy from Descartes to Nietzsche, on environmental philosophy and ethics, and on the philosophy of film. He is the founder and director of "International Cinema at Eckerd College," a successful film series for the Tampa Bay region, and is co-director of "Visions of Nature, Voices of Nature," an annual Environmental Film Festival. He completed his Ph.D. at the Pennsylvania State University in 2000. andersnt@eckerd.edu

Robert Bass is an Assistant Professor of Philosophy at Coastal Carolina University in Conway, SC. Before coming to Coastal, he taught at the University of North Florida and for Nova Southeastern University. His research and teaching interests include ethics and decision theory, and he has a long-standing interest in philosophy of religion. rbass@coastal.edu

Simon Evnine is an Assistant Professor of Philosophy at the University of Miami. His areas of specialization are epistemology, metaphysics, and the philosophy of mind. His book *Epistemic Dimensions of Personhood* is forthcoming from Oxford University Press. sevnine@miami.edu

Joseph Grcic is a Professor of Philosophy at Indiana State University. His published books include *Ethics and Political Theory, Facing Reality, Logic and Life, Moral Choices and Rawls and the Constitution*, forthcoming. jgrcic@isugw.indstate.edu

Scott Kimbrough is an Associate Professor of Philosophy at Jacksonville University in Jacksonville, Florida. His main research interests include philosophy of emotion, early modern philosophy, and philosophy of language. skimbro@ju.edu

Kirk Ludwig is Professor of Philosophy at the University of Florida. He works primarily in the philosophy of language, philosophical logic, philosophy of mind and action, and epistemology. He has published many articles in these areas and two books with Ernie Lepore, *Donald Davidson: Meaning, Truth, Language and Reality* (OUP 2005) and *Donald Davidson's Truth-theoretic Semantics* (OUP 2007). He is editor of *Donald Davidson* in the Contemporary Philosophy in Focus series (CUP 2003). kludwig@phil.ufl.edu

Piers Rawling completed his Ph.D. in 1989 in the Group in Logic and the Methodology of Science at UC Berkeley under the direction of Donald Davidson. He is now Professor and Chair of Philosophy at the Florida State University. He has wide-ranging interests and has published articles and book chapters in areas such as decision theory, ethics (with David McNaughton), the philosophy of language, logic (with Stephen Selesnick), and the philosophy of science. prawling@mailer.fsu.edu

James E. Roper joined the Michigan State University Philosophy Department after graduate work at Princeton. His teaching areas include social and political philosophy, philosophy of science, logic, business ethics, government ethics, medical ethics, and the philosophy of sport. He founded MSU's National Champion Debate Team and Debate Institutes. His articles have appeared in *Philosophy of Science*, *Synthese: An International Journal for Epistemology, Methodology, and Philosophy of Science*, *The Journal of Business Ethics*, *The International Journal of Ethics*, *American Philosophical Association Teaching Philosophy Newsletter, Florida Philosophical Review*, *Essays in Philosophy* (for which he edited the June 2005 issue on business ethics), *In the Socratic Tradition*, *Cross-Examination Debate Association 1991: 20th Anniversary Conference Proceedings*, *Research in Ethical Issues in Organizations*, *Midwest Philosophy of Education Society Proceedings*, and two articles in *Research in Ethical Issues in Organizations*. He has also published many newspaper and magazine articles. Several articles are forthcoming, including entries in *The Encyclopedia of Business Ethics and Society* and two papers in *Midwest Philosophy of Education Society Proceedings*. roper@msu.edu

Mindi Torrey is a student at the University of South Florida seeking a degree in Philosophy with a minor in Women's Studies. She intends to pursue a Ph.D. in Philosophy following graduation in the spring of 2008. mtorrey@mail.usf.edu

Jesse Unruh received his BA in Philosophy and Religion at Flagler College in St. Augustine Florida, and his M.A. in Philosophy at the University of Bristol in Bristol, England. His research interests focus mainly on contemporary practical ethics, metaethics, and moral epistemology. He intends to pursue Ph.D. studies in the fall of 2008. jiunruh@yahoo.com

Chris Zarpentine is a graduate student at the Florida State University, where he recently received his Masters in Philosophy. His philosophical interests are broad, but his research focuses mostly on the intersection of empirical science and philosophy, especially issues in moral psychology. czarpent@mailer.fsu.edu