The Yale PHILOSOPHY REVIEW

An Undergraduate Publication

ISSUE 4 | 2008

The Yale Philosophy Review is an annual journal that showcases the best and most original of philosophic thought by undergraduate students, worldwide. The goal of the *Review* is to promote discourse of the highest standard and to bring together a community of young philosophers both in the United States and abroad. Each issue contains a selection of essays on a broad range of topics as well as interviews with notable contemporary philosophers.

ISSUE IV, 2008

www.yale.edu/ypr
yalephilosophy@gmail.com

EDITORS' NOTE

Keep quiet, and people will think you a philosopher.

—Latin proverb

Brian Earp & Annabel Chang,
Editors-in-Chief
The Yale Philosophy Review

The Yale Philosophy Review

Issue IV, 2008

CONTENTS

The Surprise Examination Paradox: a Rejection of Quine, and Alternate Solutions

MATTHEW J. KNAUFF
Princeton University

In this paper Matthew J. Knauff examines one of the four primary epistemic paradoxes set forth by Jonathan Kvanvig—the surprise examination paradox. He begins by offering a statement of the paradox, after which he considers a solution proposed by W.V.O. Quine. Knauff argues that Quine's solution to the paradox must, for a number of reasons, be rejected. Finally, he offers a resolution to the paradox by means of his own approach.

In the so-called surprise examination paradox (SEP), one of the four primary epistemic paradoxes set forth by Jonathan Kvanvig, a teacher announces to her class that there will be a surprise quiz the following week. The students reason to themselves that the teacher cannot administer the quiz on Friday, for if she does then the students will know, at the end of Thursday's school day, to expect a quiz the following day—in which case the quiz will not be a surprise. The students apply the same reasoning to Thursday: at the end of Wednesday's school day, the students know that the quiz must be the next day, because Friday has been ruled out. The students continue to eliminate the days in this way until they conclude that it is impossible for the teacher to administer a surprise quiz. To their dismay, the students receive a quiz on Wednesday, and are completely surprised. Hence the paradox: the students declared a surprise quiz to be impossible, but nevertheless the quiz occurred and the students were surprised.

I say "so-called" SEP not because that is how it is commonly termed, but because I believe that the SEP is improperly labeled a paradox; as Quine says in an article on the topic, "[t]here is a false notion abroad that actual paradox is involved" (65). The surface reasoning of the SEP appears slick enough, but nevertheless one should come away from the problem with one eyebrow raised. In this paper, I will offer a close examination of the problem's construction, showing where it goes wrong. But before doing so, I look at a broader solution to the problem, formulated by Quine.

Quine considers a case that is structurally similar to the SEP: the case of the condemned man who is sentenced to be hanged at noon on some day in the next week, where it is decreed that he will be kept ignorant of the day. The former

part I call the "hanging dictum," the latter I call the "ignorance decree." The condemned man reasons in the same way as the students and concludes that his sentence cannot be carried forth without violation of the ignorance decree, and that therefore it will not occur. Unfortunately, the hangman's knock the next day proves that somewhere the condemned man's reasoning has led him astray.

Quine approaches the problem with a skeptic's scythe. It is a mistake, he says, to believe that the announcement of a surprise future event warrants beliefs about the future. In other words, Quine asserts that the conditional

> (QC) If I am sentenced to be hanged at noon on some day in the next week, then I will be hanged at noon on some day in the next week

is not necessarily true. This reasoning allows Quine to escape the paradox by introducing two more options to the condemned man's reasoning:

> (QP1) The hanging will not occur on the last day or any day prior;
> (QP2) The hanging will occur on the last day, and the condemned man will remain ignorant all the while of this possibility.

Note that the possibility of QC's being false allows for the possibility of QP1's being true. If it is possible that one is sentenced to hang but is never actually hanged, then it is consistent for the condemned man to believe in the possibility that he will not be hanged in the next week. Furthermore, QP2 hinges on this result. Because the condemned man does not necessarily believe that he will be hanged at all, it does not follow that, on the second-to-last day, he is certain that he will be hanged the next day. From this follows QP2; the hanging may take place on the last day without the ignorance decree being violated, for being hanged at all may remain contrary to the man's expectations. Furthermore, since a hanging on the last day would no longer violate the ignorance decree, the possibility that the hanging will take place on the last day can no longer be excluded, and no regressive reasoning can get off the ground, and no paradox arises.

Given this, Quine sets out the problem as follows. As we have seen, the condemned man reasoned out just two possibilities for the second-to-last day: either (1) the event will have occurred at or prior to that time, or (2) the event will occur on the last day. The latter possibility violates the ignorance decree, so he chooses (1), from which the regressive reasoning arises. By the above reasoning, though, the man should have acknowledged two additional possibilities: (3) the event will not occur by the end of the week, and (4) the event will occur on the last day, and the condemned man will "remain ignorant

meanwhile of that eventuality (*not knowing whether the decree will be fulfilled or not*) [italics added]" (Quine 66). The former option violates the hanging dictum, but the latter option satisfies each of the dicta.

To illustrate the value of this line of reasoning, Quine considers a new case, wherein the number of days on which the hanging may take place is just one, rather than, say, five, as in the SEP. Suppose a judge sentences a man to be hanged the next day, where it is decreed that the man is to be kept ignorant of the day. The condemned man will think that the judge is contradicting himself—it certainly seems strange to say to a person, "you will be hanged tomorrow, and you don't know on what day the hanging will take place"—but the arrival of the hangman, in Quine's words, shows "that the judge had said nothing more self-contradictory than the simple truth." The man should have reasoned out the following four possibilities: (1) he will be hanged tomorrow and will know it, (2) he will not be hanged tomorrow and will know it, (3) he will not be hanged tomorrow and will not know it, or (4) he will be hanged tomorrow and will not know it. The first three may be rejected on the grounds that they violate the sentence; the first violates the ignorance decree, the third violates the hanging dictum, and the second violates both. But the last possibility—which, note, is identical to QP2—is consistent with the judge's sentencing. So "[r]ather than charging the judge with self-contradiction," the man should "suspend judgment and hope for the best" (Quine 67).

Technically, this approach resolves the paradox. If the condemned man believes in the possibility of QC's being false, then he may believe in the possibility of QP1's being true, from which he may come to believe in the possibility of QP2's being true. If he believes that QP2 is *possible*, the hanging can be executed without paradox. Note, however, that he cannot actually *believe* QP2; if he believes the first conjunct, he cannot believe the second, and vice versa. To illustrate this point, Quine considers a mathematician who assumes that Fermat's Last Theorem is true "for the sake of exploring the consequences" (66). But the mathematician is not assuming that he *knows* that the theorem is true. The difference between the two is that "the latter would actually be a contrary-to-fact hypothesis, whereas the former may or may not be" (66). The same line of thought applies to the condemned man's position in regard to QP2.

But is it legitimate to think that the condemned man would actually believe in the possibility of QP2's being true? That is, while Quine's argument is valid provided that one believes in this possibility, would one ever believe it in practice? Is his solution generally true?

In general, I think the answer to these questions is "no," for a number of reasons. First, Quine's argument does not acknowledge that the word "ignorant" is subject-sensitive: whether the truth conditions for "ignorance" are met

depends on the subject in question. In other words, it is legitimate for a condemned man to believe, given the announcement of his hanging, that he is *not* ignorant of the fact that he will be hanged, in which case Quine's solution does not succeed. The one-day example of the hangman problem illustrates the legitimacy of this belief. The intuition is to say that the condemned man knows that he will be hanged the next day. Quine, on the other hand, dresses a skeptic in prisoner's clothes, and in doing so "solves" the paradox. Because the skeptic claims to be ignorant of future facts (except perhaps a priori truths; in any case, c.f. "will" in the consequent of QC), the ignorance decree is always satisfied. While the solution is valid if one uses the implausibly easy-to-satisfy truth conditions set forth for "ignorance," generally speaking—or better yet, speaking in ordinary language—the approach is untenable.

The result becomes even more dissatisfying when we realize that Quine not only might be using too-easily satisfied truth conditions for "ignorance," but that he may also be using the term in a sense altogether inappropriate to the problem. Closer examination of the ignorance decree reveals that what is meant is something like the following:

> *P*: Person X knows that event Y will occur.
> *Q*: X knows when Y will occur.

where the ignorance decree is satisfied if and only if $P \wedge \neg Q$. But Quine argues the following:

> QC2: If $\neg P$, then $\neg Q$ (for the obvious reason that if a person does not know that an event will occur at all, he does not know when it will occur).

Quine arrives at $\neg P$ by virtue of his skepticism, and therefore arrives at $\neg Q$. Instead of $P \wedge \neg Q$, Quine finds himself with $\neg P \wedge \neg Q$. This detail initially goes unnoticed because the only relevant part of the ignorance decree seems to be $\neg Q$, the statement that X does not know when Y will occur. But this is incorrect: one rightly intuits that "surprise," when used in a sense relevant to the paradox, involves ignorance with regard to the time at which the event occurs, while presupposing that the event will occur. Quine's solution does not allow for this reading, counter to our intuitions. He might argue that the paradox acts as a reductio on these intuitions, proving that they are false and that we should believe in the possibility that QC is false; after all, doing so resolves the paradox. So, as noted, while Quine's solution is viable if we allow for the

possibility of $\neg P \wedge \neg Q$, it seems to me that this approach evades the problem, forcing us to relinquish the very intuition that makes the puzzle interesting and give into skepticism. Rather than caving to skepticism, we should see if there is a solution that allows for the intuition that $P \wedge \neg Q$ is possible.

Additionally, moving back somewhat, one might think that if Quine can offer objective ascriptions of ignorance, then his solution will work even if the condemned man believes that QC is true. In other words, if he can demonstrate that we are all ignorant of future truths despite our intuitions, then he can show that his solution is generally viable. But whatever arguments he offers in support of this, he will surely not convince many people. Although this does not prove that Quine is wrong, it does show that his position is counterintuitive—and therefore unsatisfying if one believes intuitions should count. Is there a solution to the paradox that allows us to retain our intuitions?

We now turn away from Quine. In order to get a different take on the paradox, let us return to the condemned man's initial reasoning. Why does he place faith in the inviolability of the ignorance decree to such a degree that he rejects the hanging dictum? After all, the hangman's knock proves that he was wrong to do so. To phrase it in SEP terms, why do the students believe that the teacher is so committed to the exam's being a surprise that, rather than allow the possibility of a Friday quiz, they believe she will give no quiz at all? If anything, intuition tells us that of the two dicta, exam and surprise, the teacher would more likely be willing to violate the latter.

Imagine a scenario in which the students' reasoning reflects this intuition. According to the teacher, a surprise exam is to occur; that is, it must be a surprise, and it must occur. The students reason in the familiar way, and decide that no exam will take place. But this cannot be right, they say to themselves; their teacher is sure to give an exam, even if it comes at the expense of its being a surprise. After all, if she does not give the exam, she will not be able to calculate final grades.

Next, the students reason that the teacher, who is of course much wiser than they are, must be aware of this conclusion and that therefore she'll announce the date of the exam. But this is not quite right. Just because the teacher cannot guarantee that the surprise dictum will never be violated does not mean that she will not maximize the number of days on which the exam will be a surprise. Instead, she needs only to make a small concession. Namely, the teacher may allow for the exam to occur on Friday, admitting that if it does it will not be a surprise, but that the concession stops the recursive reasoning of the paradox and allows a surprise exam to occur on any of the remaining days.

By dint of considering pragmatics and taking the issue with a grain of salt, the students reword the teacher's initial utterance to mean "you will have an

exam next week, and it will be a surprise unless it occurs on Friday." To generalize, the statement

> (S1) an event *K* will occur on one of the next *N* days, and *K* will be a surprise.

is reasonably interpreted to mean

> (S2) an event *K* will occur on one of the next *N* days, and *K* will be a surprise unless *K* occurs on the *N*th day. Therefore, the probability it will be a surprise is $\frac{N-1}{N}$.

It appears then that we have found something wrong with the students' initial reasoning. Had they reasoned as above, they would have expected an exam and expected that it would most likely be a surprise. But from this no paradox arises: it is not the case that the students disbelieve the possibility of surprise, and subsequently are in fact surprised.

The result has a few interesting consequences. First, announcing the exam more in advance rewards the teacher with a higher rate of surprise; she should announce the exam as early as possible, and it even behooves her to change the unit of time, say, from days to hours (e.g., for her to say "a surprise exam will take place on some hour over the next 120 hours"). And as for the scenario in which the announcement comes just one day in advance of the exam, the above reasoning works well. The probability the event will be a surprise is $\frac{N-1}{N} = \frac{0}{1} = 0$, which falls in line with intuition: the statement "tomorrow you will have a surprise exam" certainly seems strange.

Keep in mind that we have resolved the paradox by rendering the teacher's statement differently; rather than say that the students will, without exception, be surprised, we have said that they will *probably* be surprised. First, this should not trouble us too much, for the reason that we have an account of the paradox that retains the intuition that $P \wedge \neg Q$ is possible. Even though we are left with the possibility that the surprise decree is violated, we should favor this solution to Quine's skeptical solution.

Interestingly, if the students reason in this way, it is still possible that the teacher may give an exam that is *always* a surprise (provided that $N > 1$). Imagine a very clever teacher with slightly less clever students. The teacher knows that the students discarded their regressive reasoning and decided that they would receive a quiz. She also knows that the students reasoned as above, and decided

that she meant "the quiz will be a surprise, unless it is on Friday." She then goes one step further: As long as the students are committed to this last line of reasoning, she can always give the quiz on Monday, Tuesday, Wednesday, or Thursday, and on any of these days the quiz will be a surprise. Of course, the students might predict that the teacher has reasoned like this. If they do, they realize that the teacher has excluded Friday as a possible day, and the familiar regressive reasoning takes off. But this final thought should not worry us. In practice, surprise quizzes seem to be given in just this manner: The teacher announces the quiz, never gives it on the last possible day—despite what the students might think—and the students are surprised. In short, as long as the students believe that the teacher allows for the violation of the surprise decree, the teacher need not ever violate the decree.

Rather than attempt to resolve the paradox in this way, we can also try to find something inherently wrong with the teacher's statement. In other words, though it is true that there is no such thing as a true contradiction, a person can surely utter one. Is it possible that the SEP arises in such a fashion?

Shaw suggests this in his 1958 paper on the topic, in which he maintains that "in essence the paradox is of the familiar *self-referring* type" (382). He says the paradox arises when, in addition to the examination dictum, we take the teacher to set forth the following:

> *Rule 2**: The examination will take place on such a day that on the previous evening the pupils will not be able to deduce from *Rules 1 and 2** that the examination will take place on the morrow. (384)

Shaw says that the rule makes it clear that "the origin of the paradox lies in the self-referring nature of *Rule 2**" (384). He does not, though, suggest that there is anything inherently self-contradictory in what the teacher has said. Rather, the students have misinterpreted what the teacher means by "surprise": it should mean "not deducible from certain specified rules of the school" (382). Given this, the rules of the school should be:

> *Rule 1*: An examination will take place on one day of next term.
> *Rule 2*: The examination will be unexpected, in the sense that it will take place on such a day that on the previous evening it will not be possible for the pupils to deduce *from Rule 1* that the examination will take place on the morrow. (383)

With this, only an examination on the last day would be in violation of the school rules (it would violate Rule 2). But as Shaw points out, "*any other choice for the day of the examination would satisfy both Rule 1 and Rule 2*" (383). As long as we

adhere just to these two rules, then in an *N* day term, each of the first $N-1$ days is appropriate for the choice of a surprise exam. This account also falls in line with the intuition that there is something contradictory about the statement "tomorrow you will have a surprise exam." In this case, $N=1$ and $N-1=0$. There are zero days on which the exam may be given. That is, it is impossible to carry out the statement: either the exam is not given, or it is given on day *N*, each of which violates one of the statement's dicta.

This solution is compatible with the latter part of the pragmatic solution offered above: The teacher can announce a surprise quiz and administer it on any day but the last day. Furthermore, each of these solutions retains the intuition that $P \wedge \neg Q$ is possible. They also resolve the paradox. If one reasons in either of these ways, then one will not think that the students will expect an exam and subsequently be surprised when the exam occurs. Rather, one will think that the students will expect a surprise quiz, and that they will indeed be surprised when the quiz occurs. Finally, because each of these solutions successfully resolves the paradox and retains our intuitions that $P \wedge \neg Q$ is possible, they should be
favored over Quine's solution.

Works Cited

Kvanvig, J. "The Epistemic Paradoxes." Routledge Encyclopedia of Philosophy. Ed Edward Craig. London: Routledge, 1998.

Quine, W.V. Mind, New Series, 62.245 (1953), 65-7.

Shaw, R. Mind, New Series, 67.267 (1958), 382-4.

Contemporary Moral Theory, Personal Commitments, and the Importance of Institutions

GRAHAM RHYS GRIFFITHS
University of Washington

In this paper Graham Rhys Griffiths discusses Catherine Wilson's assertion that many contemporary moral philosophers, their professed aims notwithstanding, ultimately provide justifications for the affluent lifestyles of citizens of developed nations. Though Wilson believes that these theorists, of whom she cites Susan Wolf and Thomas Nagel as examples, raise important points regarding the value of our personal commitments and their role in enabling us to live good lives, she suggests that they diminish the real requirements of a more impartial morality. Griffiths argues that this claim, as Wilson applies it to Nagel, is unfair. First, Griffiths shows how Nagel's relaxation of an impartial morality's requirements constitutes not a justification of our current lifestyles, but a dispensation due to our weaknesses. Second, he argues that Nagel does not go as far as Wilson herself in accepting the centrality of personal commitments to our lives. Finally, Griffiths argues that Nagel's emphasis on institutions that alleviate gross inequalities gives a practical approach to creating a world in which we can all live up to the demands morality makes on us.

Introduction

Philosophers are notorious for priding themselves on challenging conventional beliefs and following arguments to their logical conclusions—regardless of the political, religious, or cultural implications. Catherine Wilson, however, worries that current moral philosophers have abandoned their "obligation to criticize" in favor of lowering the threshold of moral goodness to accommodate a certain lifestyle, namely that enjoyed by citizens of developed countries. As evidence, she points to the arguments of philosophers Thomas Nagel and Susan Wolf. At the same time, Wilson does believe that Wolf, in particular, raises important issues concerning the importance of personal pursuits in the individual's life, and the inability of impartial moral theories to justify these pursuits. Given her desire then to accommodate Wolf's insights but resist the philosophical trend of which she believes Wolf is a part, it seems strange that Wilson should single out Nagel's work as an object of criticism. I will argue that Nagel's work does not constitute part of the trend Wilson rejects and that

further, it does not accommodate all of the points from Wolf that Wilson accepts. Accordingly, I will first elucidate Wilson's observations regarding current moral philosophy before proceeding to examine Nagel's work in light of these observations. In the course of this examination I will consider a second criticism Wilson directs at Nagel before replying to it.

Wilson's Argument

Wilson notes that several prominent works of contemporary moral philosophy take as a point of departure the question of whether the consumption and leisure habits of the typical middle-class Westerner can be morally justified. Wolf proffers perhaps the most extreme answer in her essay "Moral Saints." In it, she argues that morality has nothing authoritative to say regarding the pursuit of these activities because morality, as Wilson paraphrases her, "is not the only value, the one to which every other good must be sacrificed" (279). In fact, subjecting all our actions to a moral litmus test would have a dehumanizing effect on us, because it would prevent the development of other equally important aspects of our lives.

Wolf's essay, Wilson argues, can be seen as a response to the revisionist theories[1] of philosophers like John Rawls and Peter Singer. Both of these philosophers impose stringent guidelines of right action on individuals. For Singer, the fact of gross inequality in the world obligates all relatively wealthy individuals to give away substantial portions of their income. If, for example, Lisa wants to buy a plane ticket to visit her parents, but knows that money would produce more good were it donated to famine relief in Niger, then Lisa should forgo the plane ticket and donate the money. Lisa would then be required to subject every aspect of her life to this moral calculus. The problem with Singer's theory, as Wilson notes, is that it confronts "the limits of philosophy as a discursive mode," because it cannot "bring about the effect in the reader that its content mandate[s]" (280). This failure stems from its total disregard for the importance that the commitments and pursuits it commands us to sacrifice play in our lives.

Given the motivational impotency of these stringent moral theories, Wilson finds it natural that philosophers should challenge the authority of moral philosophy by asserting the importance of the particular and the personal in our lives. Wolf, again as the extreme case, argues that "no philosophical theory can tell us how important moral goodness is against other forms of goodness" (Wilson 280). Wilson applauds these challenges:

[1] i.e., theories that require us to revise our values or the way we live.

> By reminding us of what we do care about and what the effect of not caring about these things would probably be on our lives, they show that we do not and cannot in fact subject every action to moral scrutiny, and suggest that the question of whether we should do so is otiose. (Wilson 284)

That is, arguments like Wolf's draw attention to the major failing of previous moral philosophers.

Nevertheless, Wilson sees in the work of Wolf and Nagel the seed of something pernicious. In upholding the importance of individual commitments and projects to our lives, Wolf and Nagel invoke the importance of literature, tennis, the theater, fine dining, glassware, and classical music. They then go on to ask whether moral theorists can really demand that we surrender what gives shape and meaning to our lives and provides us with pleasure. But, Wilson notes, all of these pursuits "are plainly bound to a[n] economic context and a milieu that is established in the large or middle-sized cities of the Western industrial democracies" (286). To analyze this seemingly trivial observation, Wilson references the work of Karl Manheim. His theories hypothesize "that the intelligentsia of a society, especially its philosophers, are the people whose task it is to produce an interpretation of life for that society" (277). According to this theory, what appears to be the new moral philosophers'[2] assertion of the importance of particular projects and commitments is in fact a defense of the importance of the projects and commitments of a particular class. Rather than exposing the limits of morality, therefore, Wilson argues the work of the new moral philosophers "constitutes a defense of a life of leisure and privilege" (284). In order to conceal this, part of the defense "involves the inclusion of a specific statement to the effect that it is not a defense" (284).

Wilson believes this is a potentially serious charge. For, she argues, "implicit in the practice of philosophy is the recognition of an obligation to criticize that which one is naturally inclined to believe and to do" (288). In rejecting the revisionist ethics of Singer and Rawls, the new moral philosophers have also abdicated part of their responsibility as moral theorists—to criticize the values of their culture from a detached perspective. Wilson asserts that this development should be seen as being at least as deleterious as the failure of revisionist theories to understand peoples' motivations and the importance of personal projects and commitments.

[2] In this paper I use the term "new moral philosophers" to refer to philosophers, such as Wolf and (supposedly) Nagel, who attempt to accommodate the importance of the personal and the particular.

What Wilson finds particularly frustrating about the new moral philosophers' failure in this regard is that she believes that adopting the view that particular interests and commitments matter does not commit one to a theory that treats the pursuits of affluent Westerners as without need for justification. Rather, she argues, moral philosophers must incorporate the critique of revisionist theories into their work while still providing the tools with which to evaluate moral worth. While acknowledging the importance of the particular in the composition of our lives, and therefore the limit of the demands that morality can make on us, moral philosophers need to provide the criteria for determining whether we must call "people who excel at music, gardening, entertaining, the appreciation of literature, or at study, or writing *good* [italics added]" (Wilson 287). Wilson's claim against contemporary moral theorists, then, is that they abdicate their philosophical responsibility in favor of providing covert justifications of their own lifestyle.

A Defense of Nagel

Given the seriousness of Wilson's charge against Nagel, we must ask whether it is justified. My argument is that it is not. Far from abandoning his role as a moral philosopher in order to justify the lifestyles of the (relatively) affluent, Nagel, while acknowledging its critics, does preserve the authority of impartial morality. In order to demonstrate this, I will attempt to briefly lay out Nagel's project. He first to examines the possibility of conflict between the good life and the moral life. He argues "that it is the task of a moral theory to tell us not only what we are morally required to do but also how to lead a good life," and it is regarding the latter part of this task, he argues, that the revisionist theories fail (Nagel 195). Suggesting that the resolution of this conflict is the major problem facing moral philosophy, Nagel considers five positions moral theorists could take to resolve it.

First, one could argue that the moral life is defined in terms of the good life. This argument would subject all moral requirements to the test of whether they promoted a good life. Second, and conversely, one could argue that the good life is defined in terms of the moral life. This argument would make the goodness or badness of a life conditional solely on whether that life was moral. Nagel believes both of these arguments make fundamental mistakes. In regard to the first position, he states that because "moral requirements have their source in the claims of other persons" there is no reason to assume that these requirements could always be accommodated within the individual good life (197). The second position, on the other hand, ignores the complexity of our lives by claiming that they can be judged as good or bad on the basis of a single criterion.

Third, one could argue that the good life overrides the moral life. Unlike the first position, this view defines the moral life and the good life independently, but argues that an individual should act on moral reasons solely to the extent that these actions contribute to a good life. This position, however, ignores the deep-seated need we have "that personal projects and individual actions can be harmonized with universal requirements … typically expressed by certain moralities" (198). That is, Nagel believes we find something valuable and fulfilling in adopting the moral point of view, from which we regard ourselves and our personal projects as no more important than others and their projects. A life in which this point of view is ignored in favor of simply pursuing one's own ends will be one in which something deeply important is disregarded. Moral reasons will therefore often outweigh the importance of pursuing a good life.

Nagel believes that it is between the two remaining positions that the real debate exists. The fourth position states that the moral life overrides the good life, while the fifth position states that neither the good life nor the moral life consistently overrides the other. Nagel himself supports the fourth position, "that the correct morality will always have the preponderance of reasons on its side" (199). For this to be true, however, he must show that we always have reason to choose the moral life when we are confronted by a conflict between it and the good life. That is, the moral life must always be consistent with the rational life.

He considers two interpretations of what could be meant by claiming that acting morally is consistent with acting rationally. One might mean either that it is always irrational to act immorally, or one might mean that it is never irrational to act morally (Nagel 200). The truth of the former claim would ensure the truth of the fourth position and is thus the stronger of the two claims. The truth of the latter claim, though weaker, would at least point to the correctness of the fourth position. Nagel believes he can prove at least one, if not both, of these claims.

This, then, represents the critical juncture in Nagel's work. He wants to advance the claims that the moral life overrides the good life and that the moral life is, at least in a weak sense, rational. Theorists such as Rawls and Singer accomplish this task by devising moral theories that require rational individuals to subject their lives to impersonal moral scrutiny and to change their lives in accordance with this criticism regardless of its impact on the quality of their lives. Wolf, as we have seen, criticizes these theories for ignoring the importance of the particular aspect of our lives. In essence, she argues that it is not rational to subject our lives to the thorough moral scrutiny of the kind advocated by Singer and Rawls. Nagel accepts this criticism and believes the result must be "a modification of the impersonal demands of morality" in order to "[reduce], and

perhaps even [eliminate], the gap between what is morally required and what is rationally required" (200).

He accomplishes this by arguing that "valid moral requirements must take account of the common motivational capacities of the individuals to whom they apply" (200-1). That is, once we have deliberated rationally and from an impartial perspective about what morality requires of us, we must modify these demands by considering, again from an impartial perspective, the limits of human motivations. The resulting modified demands, Nagel argues, will recognize "that it is unreasonable to expect people in general to sacrifice themselves and those to whom they have close personal ties to the general good" (202). Incorporating the importance of the particular into our impartial deliberations, we grant "everyone a dispensation for a certain degree of partiality" (202).

Nagel believes this reduces the possibility of conflict between the good life and the moral life by recognizing the importance of individuals' commitments and relaxing moral demands accordingly. He also believes this solution preserves the requirement that acting morally at least never be irrational, because modifying moral demands in light of the importance of the particular "does not necessarily mean that it would be irrational for someone who can do so to accept [the more stringent] demands" (203). Doing so would constitute what Nagel calls supererogatory virtue, "acts of exceptional self-sacrifice for the benefit of others" (203).

We are now in position to evaluate whether Wilson has fairly criticized Nagel. She claims that he abandons his duty to criticize the practices of his society in favor of providing a defense of a lifestyle of relative affluence. The source of this criticism has to be in Nagel's relaxation of moral requirements—in his providing us with a dispensation to devote ourselves to our personal concerns and in his making self-sacrifice in accordance with stringent moral norms a supererogatory virtue.

Seen in light of his argument as a whole, however, this does not appear to justify Wilson's criticism. Nagel argues that the moral life overrides the good life, and the *dispensation* he grants does not negate this. Rather, the relaxation of moral requirements constitutes a form of "tolerance" (204). Granting this tolerance does not mean that Nagel abandons his responsibilities as a moral philosopher: tolerance of human weakness is not rationalization of egoism. In fact, it seems quite reasonable that Nagel's theory would morally require us to sacrifice at least some of our personal pursuits, because while we may be granted a dispensation from the requirement to fully revise our lives, surely it is not outside the bounds of our motivational capacities to do more than what we currently do. Further, by emphasizing the importance of the impartial perspective in moral deliberation, Nagel maintains an evaluative foundation from which to assess peoples' moral

worth. Wilson asks whether we must call people leading Wolf's depiction of the good life morally good. Nagel's answer, as it appears to me, is "not necessarily."

If this is the case, Wilson's criticism of Nagel rests merely on her "decoding" of his use of wineglasses, theater tickets, and fine dining as examples of important personal pursuits. This method of analysis, questionable to begin with, seems baseless once we clearly understand Nagel's position. Nagel uses the importance of these things to people (who are, admittedly, likely to be middle- to upper-class Westerners) to raise difficult questions about the demands of morality, not to assert the self-evident goodness of a particular lifestyle.

In fact, it could be claimed that Nagel does not go far enough in accommodating those aspects of Wolf's critique that Wilson accepts. Wilson believes that Wolf argues convincingly that subjecting all our commitments and pursuits to moral scrutiny is otiose, due to the central role these commitments and pursuits play in our lives. This implies that certain elements of our lives are so intrinsically important that they lie outside the domain of moral judgment. Nagel, however, believes we grant each other a dispensation from morally scrutinizing all of our actions due to our *weakness*. Rather than declaring any aspect of our lives intrinsically off-limits to moral scrutiny, Nagel argues we must acknowledge peoples' general inability to *fully* subject themselves to this type of scrutiny. This point, it seems to me, demonstrates how misplaced Wilson's criticism of Nagel is: though Nagel admits that we must revise the theories of Singer and Rawls, he hopes to preserve the authority of morality as much as possible due to his deep-seated belief that the moral life ultimately overrides the good life.

A Second Criticism of Nagel

If Nagel does not abandon his duties as a moral philosopher to criticize, does he provide a reason with which we, as individuals, can justify avoiding scrutinizing ourselves? Wilson argues that Nagel distorts the possibility of subjecting one's life to the type of philosophical reflection moral theorists ought to promote. Nagel, she claims, wants to know whether someone who leads a life "that is not obviously exploitative and egotistical by the standards of the immediate community [can] nevertheless be criticized from some more detached perspective" (Wilson 286). That is, could the life of the average middle-class Westerner be criticized from a moral perspective? His answer, according to Wilson and according to the analysis presented in this paper, is that it could be criticized. But, Wilson argues, Nagel depicts the divergence between the subjective point of view (the standpoint that values our personal projects and commitments) and the objective point of view (the standpoint of impartial morality) in such a way that it would be nearly impossible to subject our own

lives to such criticism. To do so, Nagel says, one would have to make a leap of transcendence.

Wilson believes this is Nagel's statement of the basic reaction of the current moral theorists to the revisionist ethics proposed by Singer and Rawls. We have seen Wilson's criticism of this reaction. What she finds additionally illuminating, however, is how Nagel describes the effort that would be required to subject one's own life to detached criticism. Wilson argues that by phrasing this effort in terms of an incredible personal transformation, Nagel "contributes to occluding the connections which bind the public to the private" (289). Therefore, in addition to covertly providing a justification of a certain lifestyle rather than providing the basis for subjecting that lifestyle to scrutiny, Nagel also suggests that in our own lives "no reconciliation between happiness and charitable practice can be found without a leap into the unknown" (289). Wilson, however, claims that "the scrutiny of one's own life for adherence to pecuniary and other culturally determined canons of taste" need lead neither to the view that one must alter one's lifestyle in correspondence with a revisionist theory of ethics, nor to the view that one's lifestyle is entirely justified and without need of some change (289).

A Response

Once again, we must look at Nagel's account to determine whether this criticism is justified. Though Nagel feels he accomplishes the theoretical task of reducing the conflict between the good life and the moral life, he cannot escape the fact that "the basic moral insight that objectively no one matters more than anyone else, and that this acknowledgment should be of fundamental importance to each of us … creates a conflict in the self too powerful to admit of easy resolution" (205). From a practical perspective, therefore, he does not believe that personal action—because it requires this wrenching conformation of our personal commitments with the demands of morality—constitutes the best way to achieve the harmonization of the moral life and the good life. Rather, he argues, we should work to build political and economic institutions that "arrange the world so that everyone can live a good life without doing wrong, injuring others, benefiting unfairly from their misfortune, and so forth" (Nagel 206).

Wilson may be correct, therefore, in arguing that Nagel deemphasizes the importance of individual action in creating a better world. We should not view this, however, as implying that Nagel believes we should not subject our own lives to scrutiny. Rather, as I have asserted, a perfectly plausible application of his

theory would require all of us to make at least small sacrifices in order to perform our morally required duties.

He does, however, believe that performing supererogatory actions involves a wrenching transformation, because they require us to make significant sacrifices. As a practical strategy for addressing the world's ills, therefore, he does not believe we can rely on a substantial number of people making these types of sacrifices. Nagel's characterization of devoting one's life to supererogatory action as involving personal transformation does not cast doubt on the possibility of reconciling our personal commitments with charitable practice, but rather posits that we cannot expect large numbers of people to wholly subsume their personal projects to the demands of morality.

In addition to imposing on us certain moral duties that override some of our personal projects and commitments, Nagel's analysis also commits us to building institutions that can harmonize the good life and the moral life. Far from abandoning the responsibilities of the moral philosopher, therefore, Nagel's work reinvigorates the field of moral theory by elucidating an approach that lessens the conflict between what morality demands of us and what we can reasonably be expected to do. Further, he provides us with a practical approach to achieving a world in which the fundamental moral insight that no one matters more than anyone else need not produce tension with our desire to devote ourselves to our own lives.

Works Cited

Nagel, Thomas. The View From Nowhere. Oxford: Oxford University Press, 1986.

Wilson, Catherine. "On Some Alleged Limitations to the Moral Endeavor." The Journal of Philosophy 90.6 (1993): 275-289.

Cognitive Impressions

ANDREW WONG
Washington University in St. Louis

As Wong relates, the cognitive impression was the Stoic criterion of truth. The Academic skeptics challenged this criterion in a series of arguments throughout the long history of debate between the Stoics and the Academics. In response to each Academic attack, the Stoics modified their criterion in an attempt to preserve for themselves the possibility of knowledge. In the end, the cognitive impression could not withstand the attack. The reason for this, Wong argues, is not due to the irresistibility of the Academics' arguments, but rather due to the Stoics' over-modification of their criterion. Wong argues that a version of the cognitive impression without these weakening modifications is a successful criterion of truth.

That there might really exist, as the Stoics held, a criterion of truth, "a standard of judgment which assure[s] absolute certainty" (Long and Sedley p.69), needs some justification. For the Stoics, such justification came from the assumption that nature has, in its providence, ensured a means by which rational beings can apprehend the truth (Empiricus 7.253-60, Long and Sedley 40K.6-7). Assuming that an impression revealing the truth is possible, then, the Stoic agenda asked what the requirements of such an impression would be.

Kataleptikai phantasiai are properly called "cognitive impressions" only if one takes "cognitive" in a sense that allows for the impression's quality of *seizing* or *grasping* its cause so comprehensively that the cause, and indeed each graspable aspect of the cause, becomes contained within the impression.[3] Such an impression was thought to stamp the nature of its cause with utmost accuracy upon the commanding faculty of the soul (*hegemonikon*)—the part which thinks, decides, and plans—and was the cornerstone of Stoic epistemology (Diogenes 7.54, Long and Sedley 40A.1).[4] The doctrine of the cognitive impression was not

[3] An impression is much the same for us as it was for the Stoics. If a formal definition is desired, an impression is an intentional state of the soul, ultimately caused by an external object or state of affairs (adapted from Inwood and Gerson p.404).

[4] There were, of course, exceptions. Not everyone acknowledged the cognitive impression. Notably, Chrysippus thought that sense perceptions and preconceptions were also criteria (Diogenes 7.54, Long and Sedley 40A.3). For the sake of brevity, I

a stable one, however, and years of debate with the Academic skeptics led later Stoics to amend it. The final amendment, I will argue, unnecessarily weakened the practical utility of cognitive impressions. These later Stoics underestimated the explanatory force of the theory of knowledge that their elders had laid down—a theory that had already developed a lucid means by which to answer the arguments of the Academics.

Originally, Zeno set forth two requirements for an impression to be a cognitive one:

> (1) The impression must arise from what is.
> (2) The impression must be stamped and impressed exactly in accordance with what is. (Diogenes 7.46, Long and Sedley 40C.2; Cicero 2.77-8, Long and Sedley 40D.4)

(1) says that the impression must be caused by the actual state of affairs (or by an actual object). (2) says that the impression can contain only information that is *true* of the actual state of affairs (or of the actual object).[5] Furthermore, that the impression is stamped "exactly" in accordance with what is means that it must not only be accurate, but that it must be *clear* and *distinct* (Diogenes 7.46, Long and Sedley 40C.3). (1) is a requirement of all impressions, while (2) is unique to cognitive impressions (Aetius 4.12.1-5, Long and Sedley 39B.2-4). Some philosophers, notably Sextus Empiricus, acknowledged a third requirement:[6] "Furthermore, [it must be] stamped and impressed, so that all the

---

will continue to refer to cognitive impressions as the Stoics' criterion of truth with such a reference understood as applying to *almost* all of them.

[5] My use of the word "information" should not be taken to indicate a hidden presupposition that the mechanism of impression involves objective "raw sense data" that is then transformed by the commanding faculty. Rather, I take "information" to refer to the content of the impression, which can be expressed in terms of propositions liable to have truth values. It is in this sense that the "information" can be true of or false *of* the cause of the impression. (2) is what requires that all of this information be true of the object that the impression grasps.

[6] It is unclear from the textual evidence when this third requirement was introduced or who introduced it. It does not seem to have been a Stoic response to an Academic objection, as requirements (4) and (5) were. In fact, many philosophers, both Stoic and Academic, probably assumed it was included in (2). In any case, I, like Sextus, have interpreted (2) in such a way that (3) is *not* obviously included in it, hence I have what some would call this extra requirement. Nothing in my argument depends on the separateness of this requirement or on its origin. Let the numbering of the

impressor's peculiarities are stamped on it in a craftsmanlike way" (Empiricus 7.247-52, Long and Sedley 40E.6). By this he simply means that the impression must grasp everything about its object that is graspable, i.e., it must convey *complete* information about its cause, not merely true information. Thus we add:

> (3) The impression must be stamped and impressed with all of the impressor's peculiarities.

The addition of (3) is not a trivial one. By themselves, (1) and (2) only guarantee that an impression is caused by an existing object, and that it gives true information about that object in a clear and distinct fashion. If the cognitive impression is to be the criterion of truth, it surely ought to convey the *whole* truth.

As the Academics began to voice their doubts, the Stoics made adjustments to the doctrine of the cognitive impression where they saw fit. This resulted in two further requirements being added to the three above. It was the fifth and final addition, I shall argue, that was the mistake.

Some time after the establishment of requirements (1)-(3), Arcesilaus objected that if the cognitive impression were of such a kind that there could be a false impression indiscernible from it, it would not be a suitable criterion of truth. For in that case, any given impression could be a false one, and assent would never be appropriate (Cicero 2.77-8, Long and Sedley 40D.5-7). Hence, the fourth requirement was added:

> (4) The impression must be of such a kind as could not arise from what is not. (Empiricus 7.247-52, Long and Sedley 40E.7)

The pertinent question prompting the addition of the fourth requirement is this: How can we be sure that any given impression is a *cognitive* impression?

While some interpretations of (4) answer the pertinent question, others do not. We begin with the latter. Notice that (4) is entailed by the conjunction of (2) and (3) if only actual objects have the ability to cause the kind of impression that grasps all (and only) aspects of its cause. As Long and Sedley put it: "The effect of this [fourth] addition is to insist that only real things as they really are *can* produce the clarity and distinctness characteristic of cognitive impressions" (Long and Sedley p.40). If this metaphysical claim is true, then the fulfillment of (2) and (3) logically implies the fulfillment of (4), but still nothing is said to assure

---

requirements represent their (likely) chronological implementation, as well as their conceptual fundamentality.

that the impression is, in fact, cognitive. To demonstrate: if my impression is stamped exactly in accordance with what is and with all its impressor's peculiarities, and only real things as they really are can produce impressions stamped in such a way, then my impression must be caused by something real as it really is. But how can I be sure that my impression is stamped exactly in accordance with what is and with all its impressor's peculiarities? So this interpretation of (4) fails to answer the pertinent question that the addition of (4) was initially meant to address; it still offers no means by which to determine the status of the impression.[7]

The interpretation just discussed exemplifies a recurring problem of which we will see more, that it is the tendency of the Stoics (and, as in this case, their interpreters) to create *extrinsic* requirements for the doctrine of the cognitive impression. In its attempt to eliminate the possibility of exactly similar but false impressions, the above interpretation of (4) misses the point. The desideratum for purposes of the cognitive impression is *not* to be free of misleading false impressions (though that is clearly desirable for other reasons). Instead it is to *specify the requirements* for an impression to be a cognitive one. Once this is done, the criterion of truth will have been established, and, following the Stoics' belief, every rational being will be ensured a means by which to attain true intentional states of the commanding faculty. Zeno recognized the importance of this task: "He did not attach reliability to all impressions, but only to those which have a peculiar power of revealing their objects. Since this impression is discerned just by itself, he called it 'cognitive' …" (Cicero 1.40-1, Long and Sedley 40B.2). If the criterial force of the cognitive impression is to be an intrinsic quality of it (we discern "just by itself"—just by its very *nature*—that it has the "peculiar power of revealing its object"), then all requirements for an impression to be cognitive must be intrinsic as well.[8] Indeed, if the cognitive impression is the criterion of

[7] Also notice the confused wording of the quotation from Long and Sedley: "…only *real things as they really* are can produce … [italics added]." But, in fact, any object *qua* impressor can produce *either* kind of impression (cognitive or incognitive). Even Ptolemy IV Philopator's wax pomegranates (Long and Sedley 40F) *could have* impressed upon Sphaerus "as they really were" (tasty-looking but waxy and inedible pomegranate-seeming objects). However, Sphaerus' *impression* of the wax pomegranates did not *grasp* them as they really were. Hence, (if the quotation were not otherwise misguided), it would be that the clarity and distinctness characteristic of cognitive impressions can only be due to the *impression grasping* real things as they really are.

[8] We do not count the extrinsic requirement (1) here, because we are speaking of all the requirements that differentiate cognitive impressions from all other impressions.

truth in virtue of what it is, then if it were any different (had different requirements), it would cease to be the criterion of truth. But if it stays the same, it is the criterion of truth as we have just established. Now if the requirements were extrinsic, then a change in something *other* than the cognitive impression (its requirements) could in fact result in its ceasing to be the criterion of truth. The requirements of the cognitive impression must, therefore, be intrinsic, for otherwise it is not the criterion of truth *in virtue of what it is*.

In interpreting requirement four, we should take seriously that it is a requirement that the impression *be of a certain kind*. An interpretation in which (4) is reducible to an independent metaphysical claim ("… only real things as they really are can produce …") does not acknowledge that the impression must be of a certain kind, for to be of a certain kind is to have some *intrinsic* quality. Hence, I propose an interpretation of (4) that predicates of the cognitive impression some property that qualitatively distinguishes it from all other impressions that might otherwise be indistinguishable. There are two promising options for this distinguisher: (a) that the cognitive impression *compels* the assent of the agent impressed upon, or (b) that the cognitive impression is phenomenologically introspectable, i.e., that an agent can identify the cognitive impression by its special quality, presumably that it seems peculiarly striking or self-evident.

Some time later, when philosophers of what was then the New Academy began to question the fourth requirement, they did so by questioning the ability of the mechanisms (a) and (b) above to genuinely distinguish between cognitive and incognitive impressions:

> Carneades says that he will concede the rest of it to the Stoics, but not the clause "of such a kind as could not arise from what is not." For impressions arise from what is not, as well as from what is. The fact that they are found to be equally self-evident and striking is an indication of their indiscernibility, and an indication of their being equally self-evident and striking is the fact that consequential actions are linked to [both kinds of impression]. (Empiricus 7.402-10, Long and Sedley 40H.1-2)

The observation here is simple: in some cases false impressions seem to compel assent, as when, for example, an insane man believes that his own

(1) has nothing to do with the cognitive impression's criterial force. (1) is just what makes the cognitive impression *an impression*.

children are his enemy's children and consequently murders them. If this is true, then compelling assent is not a unique quality of cognitive impressions, hence (a) should be dismissed. A similar move can be made for (b). But why should we be so quick to accept that the insane man, in this case, takes his impression to be a *cognitive* one? To put it another way, why should we think that he is compelled to assent to this impression? We need not think that he was *compelled* to assent to his impression to have acted upon it. It is not a mark of the insane to act based upon good reasons, much less to act only upon those impressions which are absolutely indubitable.

I will not push this point further, however, because there is another response to Carneades' argument, one that applies more generally: "Non-cognitive [impressions] are ones people experience when they are in abnormal states" (Empiricus 7.247-52, Long and Sedley 40E.2). Just as non-cognitive impressions are uncommon and associated with abnormal states of mind, cognitive impressions are said to be common and associated only with normal states of mind (c.f. Long and Sedley 39A.5, p.240). The cognitive impression, the Stoics maintained, was never *alleged* to function successfully as the criterion of truth in such abnormal states and circumstances as those that feature in Carneades' examples.

We have now arrived at the crucial juncture. It was at this point that some Stoics took a new line of reasoning that followed from the answer given to the Carneadean objection to the fifth requirement of the cognitive impression. "[T]he later [Stoics] added the words 'and one which has no impediment.' For there are times when a cognitive impression occurs, but it is incredible owing to the external circumstances" (Empiricus 7.253-60, Long and Sedley 40K.1). These later Stoics thought that a fifth requirement had to be added to the doctrine in order to account for cases such as abnormal mental states, and so they added:

(5) The impression must have no impediment.

Unfortunately, the establishment of this claim further opened the cognitive impression to an argument the Academics had employed earlier:

> In the case of things which are similar in shape but different objectively, it is impossible to distinguish the cognitive impression from that which is false and incognitive. For example, if I give the Stoic first one and then another of two exactly similar eggs to discriminate, will the wise man, by focusing on them, be able to say infallibly whether the egg being shown is one and the same or different? The

> same argument applies in the case of twins. (Empiricus 7.402-10, Long and Sedley 40H.4)

The Stoic response was that the sage would withhold assent if he could not judge the difference. However, given certain expertise (e.g., being a chicken farmer, or being the mother of the twins), it would, in principle, be possible to tell the difference.[9] But after the implementation of the fifth requirement, the difficulty in discerning cognitive impressions from incognitive impressions will arise for *every* object that one does not have expertise about. That is, for every cognitive impression I have, there is another incognitive impression that is no different to me unless I am an expert in the kind of object that causes that impression. How can I be sure which impression I am having? Cicero brings out the further point that the Stoics' adherence to the identity of indiscernibles does not aid this problem, for even though no two things may in fact be exactly similar, what matters is that they *appear* to be so, and thereby "deceive the sense." He goes on to say the following:

> If a single likeness has done that, it will have made everything doubtful. With that criterion removed which is the proper instrument of recognition, even if the man you are looking at is just the man you think you are looking at, you will not make the judgment with the mark you say you ought to, viz. one of a kind of which a false mark could not be." (Cicero 2.83-5, Long and Sedley 40J)

Cognitive impressions are still possible in every case requiring expert discrimination, but that the criterion of truth is merely *possible* is not enough. In nearly every case, the fifth requirement will not be met, the impediment being the agent's own lack of expertise. Even the sage will have to suspend judgment much of the time, for the amount of knowledge one has and the amount of time spent practicing are less important to becoming a sage than the coherence and stability of the knowledge one already possessed (c.f. Long and Sedley 41H). Though it is true that the requirement of expertise does not, strictly speaking, end the cognitive impression's status as the criterion of truth, it directly contradicts the Stoics' observation that the majority of impressions are cognitive. It also puts a great strain on the Stoics' providential view of nature. If the

[9] The Stoics held to the identity of indiscernibles, hence no two objects could ever be *exactly* the same. This is made intelligible by their notion that every object is "peculiarly qualified" as a unique individual (c.f. Long and Sedley 173-4).

cognitive impression is viewed as a tool given by nature that enables all rational beings to thrive in the universe (Long and Sedley 40K.6; 41B.3), it hardly seems fitting that we should be asked to suspend judgment on nearly every impression.[10]

It would have suited the Stoics and their epistemology much better to have left off the fifth requirement, for there was a ready, built-in response to the objection that the cognitive impression is sometimes "incredible owing to external circumstances." The answer is that in such circumstances, there simply *is* no criterion of truth, or at least none available to the person who finds herself in an abnormal mental state or in an abnormal situation. This response is not only a perfectly plausible one, it also harmonizes with the Stoic idea of the epistemological development of the person. Take as an example of an abnormal situation Heracles having a cognitive impression of Alcestis, who had died, but who had risen from the dead (Long and Sedley 40K.1). Heracles did not have the requisite mental perceptions necessary to form a conception of rising from the dead, and hence he was unable to conceive of the possibility that Alcestis had risen.[11] His stock of conceptions, as it were, did not allow his cognitive impression to function as a criterion of truth.[12]

That this only happens in *abnormal* circumstances or states of mind avoids the mass suspension of judgment brought on by the fifth requirement. The reason that the malfunction only occurs abnormally is that most people will have developed the conceptions called upon in everyday life. In any case, to require such conceptions is to ask for much less than expertise, as with the fifth requirement.

I have given here an outline of the development of the doctrine of the cognitive impression, and have argued for its soundness up through the fourth requirement. Although I have attempted to show the Academics to be correct in their arguments after the fifth requirement had been added, I conclude that the cognitive impression, adjusted to abolish its fifth requirement, is a promising criterion of truth.

---

[10] Notice also that this move constitutes yet another recurrence of altering the doctrine of the cognitive impression such that the requirements become extrinsically, rather than intrinsically, fulfilled. Unlike (4), however, (5) cannot be interpreted as anything but a statement of an extrinsic requirement.

[11] For more on this epistemological development of the person, see Long and Sedley 39C.

[12] Whether we would even be correct in calling his impression of Alcestis "cognitive" is perhaps a valid question, but I will not look into it further here. All that needs to be made clear is that when the commanding-faculty is in some way lacking, the mechanism of the cognitive impression as criterion of truth will not properly function.

Works Cited

Aetius. Doxographi Graeci. Reconstructed by H. Diels. Berlin: 1879. (via Long and Sedley)

Cicero, Marcus Tullius, *Academica*. Teubner series. Ed. O. Plasberg. 1922. (via Long and Sedley)

Empiricus, Sextus."Against the Professors (Adversus mathematicos)." (via Long and Sedley)

The Hellenistic Philosophers, vol. 1. Ed. A. A. Long and D. N. Sedley. Cambridge: Cambridge University Press, 1987.

Hellenistic Philosophy. 2nd ed. Ed. Brad Inwood and L. P. Gerson. Indianapolis: Hackett Publishing Company, 1997.

Laertius, Diogenes.Lives of Eminent Philosophers.Trans. R. D. Hicks. Cambridge: Harvard University Press, 1925.

Note: references to page numbers are preceded by "p."; numbers not preceded by "p." are references to section numbers.

Obligation, Rationality, and Right in Fichte's *Grundlage des Naturrechts*

MATTHEW SEAN PINES
Johns Hopkins University

What happens to our notion of political obligation when right is divorced from moral considerations? When one says that an individual's claim to her property is a right that ought not to be abridged, on what kind of nonmoral principle can one rely? According to Matthew Sean Pines, Fichte believes that he can ground such a normative prescription on a theory of natural right, deriving a set of strict political principles from a necessary metaphysical conception of a rational being. It is the main task of this paper to work through this difficult deduction, assess its validity, and discuss the general implications of its result. What is said here concerning Fichte's specific proposal for an objective theory of natural—as opposed to moral—right bears broader significance for the fundamental problem of the nature and source of obligation in the political world.

One of the foundational issues in political philosophy is the relationship between morality and political right. The problem here concerns discovering the nature of obligation in the political realm. It seems that one must postulate the existence of some kind of obligation; otherwise, there would be no explanation for the fact that individuals do interact with each other and do so in ways that, at least some of the time, show a faculty for something different than unadulterated competitive ruin. The question becomes: Does this obligation find its ground in a moral capacity of individual human beings, or rather, does this kind of behavior merely represent some rationally negotiated condition of self-interest? The source of obligation in the political world is a central question because it gives rise to a crucial concept: that of human right. What kinds of things or aspects of one's existence can one make a personal claim to and expect that claim to be respected by fellow beings? More fundamentally, if this claim is supposed to be assured in some manner, to what extent is it guaranteed and on what grounds?

One can easily confuse this distinction. When first we inquire into the meaning or nature of *right* we tend to think of it as an assertion or claim. It seems clear that a general investigation like this isn't focused on what kinds of claims *are* actually upheld or enforced, but instead aims at providing a means of

deciding what rights *should* be assured, and justifying this assurance on some rational ground. This implies that basic claims of right necessitate some kind of normative prescription, a prescription that morality seems to be uniquely suited to provide. Thus, we are led by a seemingly simple argument to believe that questions of right are normative questions and therefore fall squarely in the moral sphere. This conclusion hinges on the premise that only morality can provide the requisite ground for normative prescription. If one wishes to strictly separate the moral from the political realm, however, one must find some other way to maintain the normativity of rights. The overarching endeavor of what follows is to assess Fichte's proposal for just such a theory and to examine what it entails for real political right.

Fichte takes as the primary goal of his *Grundlage des Naturrechts* (*Foundations of Natural Right*) to reject as unintelligible the prospect of founding right on morality, and instead to formulate a clear theory of objective political obligation. He wants to give a deduction from a certain set of metaphysical first principles to conditions necessary for the existence of right. If we take Fichte's oft repeated maxim that nothing grounded can extend further than its ground, we are led to examine the first steps of this deduction. Here, in these first sections the crucial moves are made that will determine the overall picture of right and the foundations for a political society.

Fichte believes that "the concept of right should be an original concept of pure reason" (9). Therefore, it must be deduced rigorously from a pure and necessary metaphysical conception of the human rational subject. One can do "real philosophy" (Fichte 7) only if one has already come to the necessary realization that the idea of a rational being must comprise the primary ground for further reasoning. We must philosophically abstract from our own particular "I" and attempt to form a conception of the general necessary features that define this being as a metaphysical entity. As such, our conceiving of ourselves as a rational being must not look any further than its immediate necessary nature. We know, as the first necessary assumption for any possible metaphysics, that rational beings (namely ourselves) actually exist, and that by examining philosophically (i.e., by means of rational introspection) the necessary features of this "I," we come to realize that it "must act in this way if it is to exist as a rational being at all" (Fichte 5).

There are, however, two aspects to the acting of the rational being and together, these two aspects constitute the whole of all being. The first is the internal self-consciousness of the rational being—acting as it does and through itself—as the self-positing of its own existence. The second is the object of the consciousness as an external constraint. Significantly, this object ("thing") is not to be conceived as an independently existing entity with a foundation of being all

its own, but as an *emergent* feature of the rational being that arises "in this acting and through this acting (*simply and solely through this acting*)" (Fichte 5).

There is thus a tension between the conception of self-consciousness as the ultimate ground of being and the conditioning of this ground by objective externality.[13] Fichte's resolution of this tension consists of a synthesis between these two and becomes his first principle of transcendental self-consciousness—that of the self-positing "I." This self-consciousness is the ultimate ground of experience and also the ultimate ground of objective externality.[1] The rest of the deduction consists in discovering and establishing conditions for the *possibility* of the realization of this self-consciousness. Each step reveals new conditions, which themselves entail further consequences for human interaction. Much of the work pushing the deduction forward comes from clarifying the concrete relevance of merely formal concepts, and determining their real application with respect to the conditions of real being. Eventually, we will see a conception of right emerge as a relation between mutually interdependent rational beings.

An important corollary of this first principle is the demand for self-determination that it implies. The "I," in its self-positing self-determination, makes the world a product of itself and its nature is imprinted throughout. The rationality of the "I" is then carried over to the world as a whole. Thus for Fichte, the world, as a product of the rational "I," is a fully rational and logically coherent structure. This is an important idea to keep in mind when we get to Fichte's later claims about the stability of his theory of right even in the face of universal selfishness. Given his strict separation between morality and right, we shouldn't be surprised if his system, constructed on a conception of the human individual as a purely rational being, can handle moral failing. What we will have to see, however, is whether his theory can deal with the seemingly irrational aspects of human behavior and belief. While the separation between morality and right might have immunized his theory from the consequences of human moral imperfection, Fichte's system might now be left vulnerable to the empirical facts of human irrationality.

The next crucial stage of the deduction begins by considering the logical consequences of conceiving of the rational being as finite, specifically, finite with respect to the exercise of its free efficacy. Thus, the concept of freedom deduced here is the unfettered *efficacy* of the rational being. It is distinguished from and opposed to the concept of a limit or opposition—the not-"I." The subject—the rational "I"—is free insofar, not as it is self-determining, but as it is "determined to be self-active by means of an external check [*Anstoß*]" (Fichte 32). This is

[13] As a formula: $I_{subject} = I_{object} \rightarrow$ self-positing "I."

clarified by conceiving of the sensation of external constraint not as a directed force opposed to the will of the subject, but as a result of the subject's ceaseless outward activity running up against an "inert, wholly passive" obstruction. Where this resistance is encountered, in the case of reciprocally interacting individuals, one demarks the boundary of the subject's "sphere of freedom." Within his sphere freedom, Fichte tells us:

> The subject has freely chosen; it has absolutely given to itself the nearest limiting determination of its own activity; and the ground of this latter determination of the subject's efficacy lies entirely *within the subject alone*. Only in this way can the subject posit itself as an absolutely free being, as the sole ground of something; only in this way can it separate itself completely from the free being outside it and ascribe its efficacy to itself alone. (40)

It is clear that Fichte wants the rational subject to be, within his sphere of freedom, completely unconstrained from any external influence—that is, any influence that does not issue from the ground of his own personal "I"—and thus, the "concept of the rational being's efficacy is construed by means of absolute freedom" (Fichte 27). Any choice, if it is to be a free choice, must be made from only this ground.

The first stage of this deduction is an attempt to deduce objectivity from self-consciousness. Here it is important to keep in mind, as Fichte tells us, the distinction between the standpoint of philosophy and that of ordinary life. Whereas the philosopher stands on transcendental ground and forms concepts of being in an abstract ideal sense, the ordinary willing agent, fully situated within the natural and practical world, forms his concepts in a raw realist manner. The advance that Fichte wants to give to this division is his claim that philosophy can, and should, explain and justify the concepts of ordinary life. While attempting to create a synthesis between "realism" and "idealism," Fichte believes that the advantage goes to the philosopher for one simple reason: he can explain freedom. While independent for practical purposes, the standpoint of life is only comprehensible from the standpoint of philosophy, and receives its grounding and meaning from the philosophical perspective.

This distinction finds a use in giving a perspicuous view of the self-determination of self-consciousness. The philosopher takes the phenomenological manifold of sensations seriously as the not-"I" and must account for its representation as objective while remaining consistent with the

unfettered self-determination of the "I." What appears as being to the self-conscious "I" must be a real, objective constraint. The importance of the philosophical perspective is that from it comes a view of the "I" as *mere acting* and reminds the "I" that such constraint cannot be truly independent, as it would contradict its basic freedom. Thus, the synthesis accomplished at the transcendental level can have a determinate parallel and practical use for the standpoint of ordinary life.

Getting back to the deduction, we have inferred the existence of basic objects outside oneself. The question then arises: What kind of objects must there be for self-consciousness to be possible? The nature of this object must do two things in virtue of the prior steps of the deduction: it must allow first, the positing of self-efficacy and second, opposition to this very positing. Thus, as the primary condition for the possibility of self-consciousness, the subject's efficacy must be

> synthetically unified with the object in one and the same moment, that the subject's efficacy is itself the object that is perceived and comprehended, and that the object is nothing other than the subject's efficacy (and thus that the two are the same). (Fichte 31)

The subject and object are putatively identical and yet must differ in some respect. This is only possible if the object is itself a subject; that is, the only possible object for an "I" is another "I." Thus, with the introduction of the other "I," we have here a preliminary conclusion of the deduction that serves as the launching pad from the metaphysical principles and their logical interrelationship into the full and unabashed realm of right. It will serve us well to examine in detail this crucial logical step.

It is a necessary feature of the intrinsic nature of the subject to be self-determining.

The synthesis of the object with the subject can only be maintained if "we think of the subject's being determined as *its being determined to be self-determining*, i.e., as a summons to the subject, calling upon it to resolve to exercise its efficacy" (Fichte 31). For this summons to have a real effect on the subject, the subject must view it as issuing from something external. Since only another free individual is capable of providing such a summons, this requires that there be another free "I" external to issue it. If a rational being is to posit itself "in response to a summons calling upon it to act freely," then it "must necessarily

posit a rational being outside itself as the cause of the summons, and thus it must posit a rational being outside itself in general" (Fichte 37).

While this may seem circular, it is not viciously so. Assuming the freedom of the rational being requires also assuming the freedom of another. By the structure of the summons, one can only get the freedom of one through the freedom of the other. Within the sphere of the individual's free choice, the individual is purely sovereign and yet, we see, must leave a portion open to receive the summons. This reflects the unanimity of individuality in Fichte's theory. The individual only has meaning, or even conditioned self-consciousness, insofar as it is related to other individuals who themselves are conditioned in the same respect.

This is expressed in Fichte's all-important doctrine of recognition. The concept of recognition is deduced from the concept of summons with the condition that two rational beings, in order to provide the summons and rationally receive it, recognize each other as rational beings like themselves and treat each other accordingly. It follows from the mutual interdependence of their freedom, and thus the interdetermination of their self-consciousness—the primary condition of all being. This focus on intersubjectivity is constructed from the notion of each individual as a *rational being*. It is decidedly

> *not* that the rational being in itself, apart from me and my consciousness, recognizes me within his own conscience (such belongs to the sphere of morality) or in the presence of others (such is a matter for the state); but *rather* that he recognizes me as a rational being in conformity with *his and my* consciousness, synthetically united in one … such that—just as surely as he wants to be regarded as a rational being—I can compel him to acknowledge that he knows that I am one as well. (Fichte 42)

With this final concept firmly in place, Fichte is in position to formulate the "relation of right" as a relation between free beings that requires that "*I must in all cases recognize the free being outside me as a free being, i.e., I must limit my freedom through the concept of the possibility of his freedom*" (49, italics mine). From my conception of my own freedom, I deduce the freedom of others and thereby deduce a limitation on myself by the principle of interdetermination. From this principle comes the conception of right as a necessary relation of free beings. All of this stems from the central project of finding conditions for the possibility of the

realization of transcendental self-consciousness—the primary condition for all being.

Now that we have the deduction of the concept of right laid out, I want to take the rest of the paper to address three distinct issues. The first concerns the assumption of rationality. The second concerns an unresolved conflict in Fichte's theory between the doctrine of recognition as a necessary condition of self-consciousness and Fichte's claim that the decision to live in a community is purely a matter of arbitrary will and cannot be coerced. Pulling the previous two together, the third strand aims at investigating exactly what kind of normativity comes from this relation of right and what consequences it has for the concept of coercion.

The principle of right is an extremely important feature of Fichte's system; in fact, his "entire theory of right rests upon it" (Fichte 42). Thus, we should take close notice of any qualification or conditions that might affect it. We explicitly encounter one such qualification where Fichte writes:

> Now I *must* necessarily treat him thus, just as *certainly* as I posit *myself* as a rational individual in opposition to him—*this is true, of course, only to the extent that I proceed rationally, i.e., with theoretical consistency* [italics from "this" my own]. (44)

A lot of the work that seems to be borne on the principle of right seems to be exercised by this assumption. Only if we assume that the agents involved act rationally does this principle hold. Further, this rational assumption is categorical: Even a single sudden irrational action dissolves the entire foundation for the relationship of right. I must be able to expect in every conceivable case that "all rational beings outside recognize me as a rational being" (Fichte 43) and treat me in accordance with that understanding.

This concept of rationality has an interesting dual status for Fichte. First, it seems to be a necessary first principle for all reasoning. Obviously, if one is attempting to logically deduce anything about human nature, one must start with some conception of the basic a priori features of that nature. If one assumes that part of human nature is irrational (i.e., inconsistent with itself), then it becomes impossible to extend any lines of logical deduction. If the ground of the deduction is inconsistent, then whatever is deduced from this ground is bound to be infected with the same inconsistency.

However, this concept of rationality seems to have another, quite different role *within* the deduction, not as part of the basis of metaphysical assumptions, but as a condition of the application of that deduction. Rationality, rather than

being some a priori assumption, becomes something one can *infer* on the basis of observed behavior: "[T]his consistency can be required and is only required for actions" (Fichte 45). Being able to physically infer the rationality of another being from the "moderation of [its] force by means of concepts" (45) lends an empirical sense to this notion that is markedly different from its status as a metaphysical first principle.

Importantly, it is this second sense of inferable rationality that is at play in the concomitant notion of expectation. Each rational being, in granting recognition of the other and thereby leaving open a sphere of freedom for himself, does so not only on the basis of his present knowledge of the other person's behavior, but, crucially, also on the assumption that the other's behavior in the past, while fully consistent with a concept of him as a rational being, can be perfectly extrapolated into the indefinite future, and that it can be guaranteed to him that this other will universally reciprocate any rationally accommodating behavior.[14] To reliably make this extrapolation, each rational being must be able to infer the rationality of the other from his past physical action.

The first issue raised by this is an epistemological one. This kind of psychological inference seems vulnerable to two points of attack. First, it seems to make the inductivist mistake in assuming that a person's past actions will resemble his future ones. Second, and more importantly, the reasoning underpinning this inference seems to be viciously circular. Each rational being, in trying to infer the rationality of his fellow beings, relies on a judgment of their physical action and its accordance with the concept of them acting rationally. To make this inference, however, requires that he assumes that such a consistent connection can exist in general between concept and action such that action of a certain kind can always be taken to imply a certain character of thought. This assumption is exactly the concept of rationality that is the goal of the inference in the first place and thus we see the same concept as premise and conclusion.

The second issue can be generalized in a troubling way by asking the following question: How does each individual in Fichte's scheme know that he is acting rationally himself? It seems absurd to suppose that he would observe his own behavior and check to make sure that such behavior didn't conflict with his concept of himself as a rational being. While we make the condition of inferring

[14] "Now just as certainly as I recognize him, i.e., treat him in the way described, so too is he with equal certainty *bound* or *obliged* by virtue of his initially problematic expression—he is required by virtue of theoretical consistency—to recognize me *categorically* and indeed to do so *in a way that is valid for both of us*, i.e., he is required to treat me as a free being." (Fichte 44)

the rationality of other beings an empirical one, it seems that we cannot apply this same condition to ourselves. The external inference requires that the rational being assess the actions of others on the basis of his concept of rationality, but he must already be assured of the soundness of this concept even to attempt the assessment.

The problem we have raised concerns the origin of this concept as each rational being applies it to himself. The sense of rationality involved here seems to be of the first type discussed above, namely, as a metaphysical first principle. Each rational being must first form a conception of his own rationality as a necessary principle that must be laid down before he can extend this concept to external beings.

This raises a concern over the status of rationality in general, as put to use within the deduction. It seems that Fichte wants to play on the dual senses of this concept. In establishing the first conditions of the "I" as a *rational* being, he is clearly employing its sense as a transcendental a priori condition whose force issues from the requirements of logical consistency. However, in laying the ground for the required expectation of mutual reciprocity that must obtain in order for the relation of right to be realized, Fichte seems to shift the ground to the second, empirically inferred sense. It is clear that each of these two distinct notions of rationality support very different kinds of claims. In not clearly separating between these two, the overall clarity of the deduction is harmed. Further, by seeming to rely heavily on the empirical sense in support of his formulation of the principle of right, Fichte's claims to strict theoretical and necessary justification are put in question.

Leaving the issue of rationality aside, I want to move to a crucial conflict that arises in Fichte's system. This conflict crops up between two claims that are supported independently, yet seem to be mutually incompatible. The first claim is a simple extension of a previous claim concerning the interdependent conditions for the possibility of the realization of each individual's self-consciousness. This recognition condition places a *necessary* condition on the possibility of free being: the self-consciousness of an individual can only be realized in the world by living in an association of mutual recognition and interaction with other rational beings. A corollary of this condition is the claim that "the human being (like all finite beings in general) becomes a human being only among human beings" (Fichte 37). The notion of a human individual is strictly nonsense outside the concept of society.

Let us contrast this claim with the following:

> Each [rational being] is bound only by the free, arbitrary decision to live in community with others,

> and if someone does not at all want to limit his free choice, then within the field of the doctrine of right, one can say nothing further against him, other than that he must then remove himself from all human community. (Fichte 12)

As Neuhouser points out, this seems to conflict with the conditions of self-consciousness:

> But if living in a political community governed by the rule of right is required, as my interpretation claims, for the self-consciousness of individuality, and if such self-consciousness is an essential part of being a person, then, contrary to some of Fichte's assertions, it no longer seems to be a matter of arbitrary choice whether or not we enter such a political order. (179)

The realization of an individual's self-consciousness was supposed to be a matter crucial to his mere existence as a rational being, and it seems inconsistent for something so categorical to be a matter of arbitrary choice. Neuhouser is "less sure" how to handle this problem other than to claim that it conflicts with other portions of Fichte's theory and is "one of the least plausible features of Fichte's theory of right" (179).

This seems to me a highly inadequate way to deal with what appears to be a significant internal inconsistency in Fichte's system and is especially dissatisfying in view of the importance of both claims for the overall doctrine of right. This inconsistency pushes us to more closely examine the conditions of each of these two principles. Looking more closely, we see the problem arises from an inconsistency over normativity. The unanimity of the individual as a condition of self-consciousness is a *necessary* condition. The rational being cannot sustain the conditions of its free efficacy *except* in the mutual interdependence of society. Yet Fichte also wants to claim that it is purely a matter of free choice, unconstrained by any external influence, whether to enter into such a society. In order to properly appreciate this apparent inconsistency, we must now turn to the third and final strand of inquiry: the source and nature of normativity in Fichte's theory of political obligation.

This entails getting back to the fundamental issue at hand of what kind of political obligation we get out of Fichte's system. The problem raised above then comes to a conflict between the deduction of the concept of right as a necessary

relation between free rational beings, and the claim that there is no categorical imperative or obligation of any kind to enter into the relationship of right with others if one does not wish to live in association with them. The crucial question we must resolve is this: What is the nature of this kind of normativity?

It certainly is not anything like a conceptual or a priori necessity. There is a clear minimal condition before the rule of right may exert its normative authority. This condition—the condition of living in a community with others—is empirical and contingent.[15] The world and the structure of human relations within it could have easily been such that no interdependence would arise. By making the condition of right contingent, Fichte seems to escape the tension one finds in moralist theories between a *categorical* conception of right and the technical and empirical factors imputed from the way humans actually associate with each other. However, while avoiding this tension, it seems that Fichte has created a new one. This new tension, as we saw above, comes from making one condition of right categorically necessary and the other condition contingent.

We have already seen the contingent originating condition of this normativity, namely the condition of community. There remains, however, the issue of what kind of obligation is at work within a functioning political society. To get a better look at this question, it is instructive to contrast this with the kind of normativity constitutive of moral judgment. Without getting too far into the dense issues and heavy baggage associated with it, let us examine the archetype of moral principles, Kant's categorical imperative. The categorical imperative is grounded on the transcendental necessary conditions of human consciousness that give rise to a certain conception of the person as a "morally autonomous" individual. Freedom, under this conception, refers to the categorical autonomy of human will. Every human possesses a genuine will that, through the unfettered exercise of reason, serves its universal duty to moral action.

Our capacity to act as moral agents stems from the absolute self-determination of our will and our bound duty to recognize and conform to universal a priori principles. Our recognition of ethical imperatives forms a set of internal obligations on our own personal will, taking the form of moral duty. These binding constraints arising from our moral constitution take the form of

[15] "The rational being is not absolutely bound by the character of rationality to will the freedom of all rational beings outside him. This proposition is the dividing line between a science of natural right and morality, and it is the distinguishing characteristic of a pure treatment of natural right. Within the sphere of morality, there is an obligation to will this. In a theory of natural right, one can only say to each person that such and such will follow from his action. Now if the person accepts this or hopes to escape it, no further argument can be brought against him." (Fichte 81)

the categorical imperative: Act as if the maxim of your actions were to form a basis for a universal law.

From the intrinsic ground of this personal moral compulsion, Kant derives a second level of obligation. When such a fundamentally moral individual enters into associations with other human agents and attempts to form a civil community, a new set of moral questions is raised and requires resolution. Borrowing the idea of a social contract founded on the (seeming) a priori principles of freedom, equality, and individuality, Kant conceives of the ideal moral society as a union of competitively moral citizens who respect the intrinsic and inalienable rights of their neighbors in the same way as they would defend their own fundamental civil liberties.

Fichte has a different take on this. Instead of relying on a moral imperative to secure political rights, Fichte lays the foundation of right on the necessary conditions of mutual recognition.[16] The conditions of intersubjectivity, however, require that each rational individual be able to expect to be treated rationally himself. He must be able to trust the other person to act in a way that does not contradict their mutual concept of each other. Fichte recognizes that this condition clearly does not hold.

A minimum condition for any theory of right like Fichte's is that it can handle the inherent egoism and selfishness of human beings. One needs to find, in Fichte's words, a law of coercion that consists of a structure of human relations operating with "*mechanical necessity* to guarantee that any action contrary to right would result in the opposite of its intended effect [italics added]" (127). The big question that Fichte then tries to address is exactly how one achieves this kind of arrangement in practice. The problem boils down to the question of the reliability or practical possibility of the antecedent in the following conditional:

> Now *if a law of coercion operates with mechanical necessity to ensure that any infringement of the other's rights becomes an infringement of my own*, then I will exercise the same care to ensure the security of the other's right as I do to ensure the security of my own [italics added] … (Fichte 129)

Fichte's answer to this problem is to attempt to show that one needs an "irresistible coercive power" (130) that can enforce and secure everyone's rights,

[16] "The source of this obligation is certainly not the moral law; rather, it is the law of thought; and what emerges here is the syllogism's practical validity" (Fichte 44).

but finds that this can only be sustained consistently "if each party were to have exactly as much power as right" (132)—which can only be attained in the civil condition of the commonwealth. From this follows the striking claim that "there can be no rightful relation between human beings except within a commonwealth and under positive laws" (Fichte 132). Under such a condition, the original problem is resolved because the "power of all … keeps each individual's power within its boundaries; and therefore there exists the most perfect equilibrium of right" (Fichte 136).

This marks a fundamental shift in the ground of right. Earlier, the rule of right was supposed to follow deductively from the mutual recognition each rational being has for all others as a condition on his own self-consciousness. Now, it seems that the rule of right finds its ground in a threat, the coercive force of the law, which is *constructed* as a material condition over society. The shift in the ground of right here seems to be from that of the *conceptual* necessity of the rationality of reciprocal beings to the *mechanical* necessity of law. It is clear that these are completely different notions of necessity. Further, it is not clear in what sense the mechanical action of law can be considered *necessary* at all.

The concept of "positive law" (Fichte 95) is introduced to denote a system of "norms" that the rational agent can survey objectively and take from the guarantee of his future safety and security. The existence and, importantly, the *completeness* of this system of norms is of crucial importance for Fichte's entire project, insofar as all *actual* judgments concerning the law of right are made *through* such a system of positive laws. Without such a system, the law of right would be a mere formal doctrine with no power to materially prescribe and sustain the necessary relations of right that must exist between rational beings in a stable society. Fichte makes some strong claims on this concept that seem to imply not just that every particular law must be justified on some rational basis, but that every law that is so justified is necessarily so. He seems to be claiming that the entire specific system of positive laws must follow *necessarily* from the nature of rational beings in general.

I don't think it is too strong a claim to say that this notion is flatly absurd. While one might grant Fichte the need to have necessary rational principles to secure the basic ground of right, it seems ridiculous to think that the application of such principles can necessarily take the form of only one particular set of positive laws. It doesn't make sense that the specific content of the law can be codified for all forms of future judgment irrespective of contingent factors. It seems perfectly consistent to maintain that although we found our basic law of right on a regulative principle of reason, we nevertheless want the flexibility to subsume new domains, new circumstances, and new modes of interaction under that framework. Fichte's treatment, insofar as it posits only one necessary set of

positive laws to hang over humanity for all time, seems to disregard some strong intuitive considerations about the elasticity of, not the grounding of law in general, but the particular form of a law in a given time and place.

It seems that Fichte missed a crucial feature of normativity in general: that "can" does not imply "ought." The whole structure of the deduction, especially the ambiguities involved with the notion of rationality and the dual grounds of necessity, seems to falter on the logical presumption that just because rational beings *can* act in a way that meets the conditions of mutual recognition, it does not imply that they *should*.

A political wrong is an act that negates the very possibility of an individual's freedom. The following question arises: How do you get from the *fact* of a political contradiction to the normative valuation that there *should not be* such a contradiction? This seems to imply that one needs something more, something like a quasi-moral principle that states "contradictions between natural expressions of freedom are wrong and should be avoided in all cases." This is something extra, because there is nothing in the basic intrinsic nature of the rational being from which a principle like this could follow. All that is contained in the concept of a rational being is an entity who can freely (i.e., with causal efficacy) determine itself and its sphere of action, and can guide this determination with concepts from its own spontaneous rational faculties.

There are only two facts here: 1) the existence of an individual in the sensible world whose actions find their sole ground in that individual's fundamental capacity as a free being, and 2) an action taken by another individual that stands in direct *contradiction* to the freedom of the first individual. We have merely the conflict, in a purely natural sense, of two natural wills. There is nothing in either of these two facts or any of the presumptions involved that could allow one to infer any normative valuation. In other words, all we have is a bare conflict of natural powers (free powers of natural individuals). There is nothing in the situation as Fichte conceives it that would entail that such a conflict is *wrong*, in the sense that it *ought necessarily* to be avoided. We naturally think of contradictions like this in normative terms, but it is clear that the logic of conflict between two natural wills cannot ground any objective normative valuation. One needs to insert this normativity from the outside. There is nothing in the natural domain, as described, from which it could logically issue.

I suppose my claim generalizes to a broad claim that there can be no objective normativity in the purely formal interactions of natural individuals, no matter how substantive their metaphysical constitution. All normative claims require a ground and a principle: the ground is the natural context, and the principle is the condition that specifies the direction of valuation (i.e., sets the scales of judgment). Fichte's conception of a rational being only supplies the

ground. The condition of mutual recognition, as Fichte himself recognizes, is problematic and cannot stand on its own. To save the doctrine, Fichte must retreat farther up the ladder of justification toward increasingly contingent rungs of support. Pushed, finally, towards the concept of coercion as the only reliable ground, the deduction—supposedly laid on a foundation of necessary implications and natural features of human rationality—falters on a confused contingency. Since this principle of coercion finds no other ground within the deduction other than the failed principle of intersubjectivity, it counts as an unjustified external premise in the deduction. And given its crucial position in the connection between the rational individual and the concept of natural right in general, it seems that the deduction as a whole loses its foundational appeal.

Works Cited

Fichte, Johann Gottlieb. Foundations of Natural Right. Cambridge: Cambridge University Press, 2000.

Neuhouser, Frederick. "Fichte and the Relationship Between Right and Morality." Fichte: Historical Contexts/Contemporary Controversies. Ed. Daniel Breazeale and Tom Rockmore. New Jersey: Humanities Press, 1994.

Problems with Gauker's Conditional Semantics

MARK ALAN WILSON
University of Nevada, Las Vegas

A significant amount of research has been dedicated to reconciling paradoxes that arise when English conditionals ("If P, then Q") are interpreted as bearing the same semantic relation as material implication in first order logic $(P \Rightarrow Q)$. For example, the statement "if it rained yesterday, then it didn't rain hard" $(R \Rightarrow \neg H)$, by the rule of contraposition, $(R \Rightarrow \neg H) \therefore (H \Rightarrow \neg R)$, should be logically equivalent to "if it rained hard yesterday, then it didn't rain" $(H \Rightarrow \neg R)$. Clearly, this would be a false utterance in English. Paradoxes similar to these have led a number of theorists to conclude that English conditionals are not truth-functional. Some have attempted to explain the semantics of conditionals in terms of situational contexts. Mark Alan Wilson examines a recent attempt by Christopher Gauker to explain the semantics of conditionals. Gauker redefines the notion of the context of an utterance and uses it to replace the notion of logical validity with contextual assertibility. Wilson argues that Gauker's notion of contextual assertibility generates at least two major problems: first, it fails on its own criteria, and second, it licenses intuitively unacceptable utterances. Further, Wilson suggests that the only way Gauker's theory might avoid these problems would be to reduce it to a mere restatement of an earlier theory of conditionals, that of Nelson Goodman.

Introduction

I examine a recent attempt by Christopher Gauker to explain the semantics of conditionals by replacing logical validity with a notion of contextual assertibility. I argue that the most straightforward interpretation of this notion leads to an undesirable disjunction: either Gauker's theory fails on its own criteria or it reduces to an earlier theory on conditionals. My argument depends upon the proper understanding of the context of a sentence in general and of conditionals in particular.

Conditionals

There is a rich body of work on conditionals centering on difficulties associated with the truth-functional analysis. Grice argues that while some

conditionals may be true by material implication, they are nonetheless misleading when hearers presume that the speaker is behaving according to conversational norms (Bennett 23). For example, take the conditional

(1) If Al Gore is president, then he is seven feet tall.

A conditional is true on the truth-functional account whenever the antecedent is false, but (1) is clearly a strange utterance in English. The problem is to figure out the connection in English between the antecedent and the consequent. Most of the work on conditionals focuses on the paradoxes that arise when English conditionals are interpreted as having the same semantic relation as truth-functional conditionals, particularly in relation to contraposition, *modus tollens*, transitivity, and antecedent-strengthening (Bennett).

Validity as Assertibility

Gauker explains conditionals by replacing logical validity with a notion of contextual assertibility (2005, 82). A conditional may still be logically valid, even while being contextually unassertible. This can account for why equating English "if … then" statements with material implication creates problems. Logical validity is not sufficient to make a conditional assertible. Gauker argues that "a conditional is assertible in the context of other formulae just in case the antecedent, together with the formulae in the context, imply the antecedent" (1987, 293). For Gauker, a context is nothing more than a set of linguistic items, or "formulae" (1988, 136-151). They could be a list of sentences, a string of symbols, a grouping of words, or even a set of other contexts. Understood in this way, language is like a board game. If you know how to play the game, then you know how to move the pieces and you understand the rules of the game. The game itself, not the world, teaches you these things. Play in this language-game resembles that of a mathematical function. A particular play of the game (a sentence or discourse) names a certain function. It will name a function from one set of linguistic items (words, sentences, etc.) to another set of linguistic items (more words, sentences, etc.). Suppose I am giving you directions on how to find my house by car when I say

(2) If you turn the corner, you'll see a red truck.

There are two types of contexts relevant to (2). First, there is the physical context of our conversation, that is, the state of affairs in the vicinity of which we are speaking (e.g., red trucks, people, or street corners) (Gauker 2003, 5). Second, there is the purely lexical context of our conversation, which is composed of the structure of our dialogue (the syntax of our utterances, the

inferences that our sentences license us to make, etc.) (Gauker 2003, 5).[17] For Gauker, the lexical context interprets "the values of a certain variable in a semantic theory" (5). An example of how this might work is anaphora. In anaphora, a reflexive noun gains its reference by being bound to an earlier noun, usually a definite description or proper name. For example:

(3) My father hates Limbaugh. Every time *he* hears *him, he* swears.

In this case, "he" refers to my father, and "him" refers to Limbaugh. We don't need recourse to conversational maxims or to shared assumptions to understand the referents of "he" and "him." They gained their reference solely on account of the structure of the discourse. The context entails a function from the elements "father" to "he" and from "Limbaugh" to "him." They were bound by a linguistic function—anaphora—which was generated solely by the lexical structure of the utterances, not by direct reference to the world.

This explains why a conditional can be both valid on the truth-functional account yet unassertible in English. Logical validity only makes it trivially assertible to say that if Al Gore is president, then he is seven feet tall. It might be assertible within the context of a logic text, or among a group of philosophers discussing logic. But in any other context it is simply unassertible, for no other context will entail a function from an antecedent like "Al Gore is president" to a consequent like "he is seven feet tall."

Gauker contends that all language, including conditionals, is understood in a similar manner. He sees his work "as a modification of Brandom's thesis" that a conditional is a codification of an inference rule (Gauker 2005, 82). Brandom argues that language is an explicit articulation of a speaker's implicit reasoning. Formal logic is a codification of the inference patterns of that implicit reasoning. Sentences are not to be analyzed according to their reference, but rather according to what role they play in inferences (Brandom 2000). Gauker modifies Brandom's thesis by replacing the notion of logical validity with a notion of contextual assertibility (2005, 82). For example, suppose that while we are talking, I pull out a match and say

(4) If I strike this match, it will light.

A sentence names a context. There is a specific lexical context relative to (4); the structure of our conversation has restricted the number of ways that this

[17] I call Gauker's contexts "sets," a term Gauker does not himself use. Gauker uses "context" to refer to many different concepts. To avoid ambiguity between *de re* contexts and the purely lexical contexts Gauker employs, I shall call Gauker's contexts "sets," since they are essentially just groupings of "words without content" (Gauker 2003).

conditional can be interpreted and the inferences that can be made from it. Within this context, according to Gauker, there are two relevant subsets, the antecedent set and the consequent set. The antecedent set might contain "I strike this match," "the sun explodes," "there is water on Mars," "Al Gore is president," etc. The same holds for the consequent set. The conditional (4) is assertible just in case there is a linguistic function in our context between the relevant antecedent element ("I strike this match") and the relevant consequent element ("it lights"). Gauker's notion of a context is unlike a possible world in that contexts needn't be either complete or consistent (1988, 136-151). These subsets of a context could contain anything, as long as they contain the elements relevant to the assertibility of the conditional (Gauker 2005, 82). The only restriction he places on them is that the antecedent set cannot contain the consequent element "it will light." In that case, I should instead assert a conjunction, i.e., "I strike this match and it will light." Gauker is motivated by an attempt to explain contexts (and semantics itself) without recourse to assumptions about speaker intentions, which he believes are untenable for formal semantics (1988, 136-151). While this conclusion is debatable, attacking it is beyond the scope of this paper.

Merits of the Theory

As functions, conditionals are not always bijections. The antecedent set of an assertible conditional often has more members than the consequent set. For example, suppose that our mutual friend Joe is in the hospital with kidney stones and I say

(5) If the nurse poisons him, then Joe dies.

By assuming (5) is assertible, we assume there is a function in our context from "the nurse poisons him" to "Joe dies." But this does not necessarily entail that there is an inverse function in our context from "Joe dies" to "the nurse poisons him." The antecedent set could contain other members (e.g., "Joe's kidney stones are actually cancer," "Joe contracts a deadly infection," or "Joe is suicidal"). If "Joe dies" is the only element in the consequent set, then there could be a function from antecedent to consequent without there being an inverse function from consequent to antecedent, for "Joe dies" would map to several elements in the antecedent set.

This is how Gauker explains the fallacious inference of affirming the consequent. It would be unassertible for me to say "Joe died, therefore Joe was poisoned." In our context there is no function from C to A just because there's a function from A to C. One of the chief merits of this account is that it can also

be used to explain contraposition and *modus tollens* without giving up those inferences. Contraposition, $P \Rightarrow Q \therefore \neg Q \Rightarrow \neg P$, has been a problem for other theories of conditionals (Bennett). For example, suppose we are debating about whether it rained last night. I note that the grass is completely dry and say

(6) If it rained last night, then it didn't rain hard. $(R \Rightarrow \neg H)$

By contraposition, (6) would also entail that

(7) If it rained hard last night, then it didn't rain. $(H \Rightarrow \neg R)$

Examples like this have led many theorists to throw out contraposition as a valid inference, despite its validity on the truth-functional account (e.g., Lewis, Bennett). But Gauker can explain contraposition in a way that allows us to keep it while explaining its strangeness in English. His explanation parallels that of denying the antecedent. There is nothing about our context that entails a bijection between the antecedent set and the consequent set. Just because there is a function from "rained last night" to "didn't rain hard" doesn't mean that there is a function from "rained hard" to "didn't rain." Indeed, these elements might not even be members of the relevant sets. *Modus tollens* is explained in the same manner.

Transitivity

Transitivity creates a problem for Gauker's theory. While transitivity, $(A \Rightarrow B),(B \Rightarrow C) \therefore (A \Rightarrow C)$, works on the truth-functional account it fails as an acceptable assertion in English. Suppose we have just eaten dinner at home and are now driving to see a movie. I cannot remember if I unplugged the iron, but I believe that I have and that our apartment hasn't burned down. I don't want to miss the movie. I say to you

(8) If the apartment burns down (A),
then I forgot to unplug the iron (B).

Seeing by the look on your face that you are unsure if I'm joking, I tell you not to worry. I know that the iron has a sensor that tells it to automatically shut off after a few minutes of inactivity. I say to you

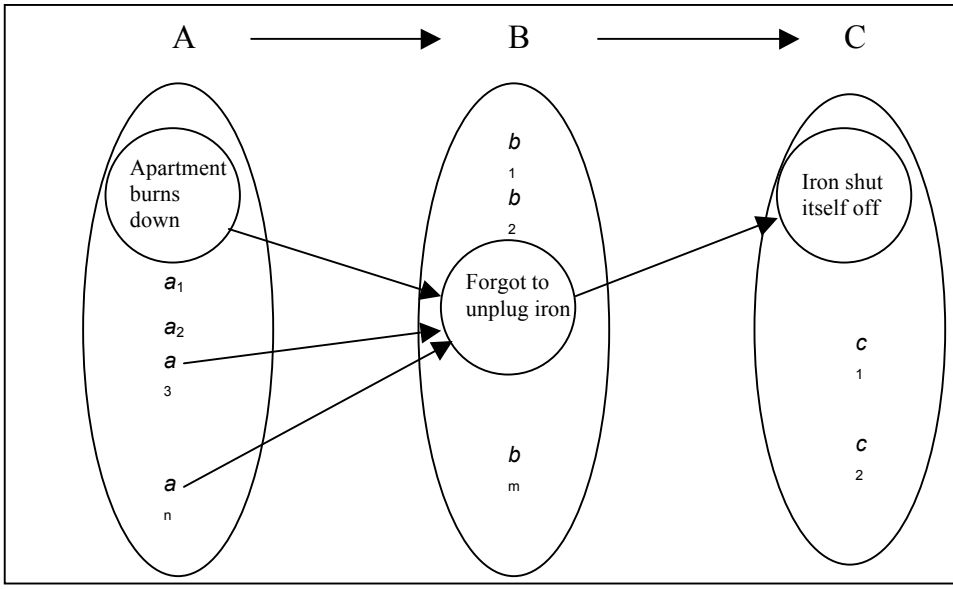
(9) If I forgot to unplug the iron (B),
the iron automatically shut off (C).

By transitivity, the first two statements entail that I should also be willing to assert

(10) If the apartment burns down (A),

the iron automatically shut off (C).

It follows from the definition of a composite function that if there's a function from A to B $(f : A \rightarrow B)$ and a function from B to C $(g : B \rightarrow C)$, then one can construct a composite function from A to C $(g \circ f : A \rightarrow C)$. In the context of our ride to the movies there is:



Gauker's concept of a language function permits transitivity. Even worse, Gauker would have to say that (10) is non-trivially assertible. It is not only a valid truth-functional inference, but it also follows from the existence of the language functions. The failure of transitivity in English is demonstrated by an example from Lycan (29):

(11) If Gore is nominated again for President, I will skip the front page of my newspaper.

(12) If all the current Democratic candidates are squashed by a meteorite, Gore will be nominated for president.

(13) Thus, if all the current Democratic candidates are squashed by a meteorite, I will skip the front page of my newspaper.

As Lycan concludes, (13) is plainly not something we would assert in English, despite its validity on the truth-functional account (29). Gauker's

context semantics, being driven by language functions, would force us to accept transitivity and the strange utterances it generates.

Vagueness

Saying that assertible conditionals have a function between their antecedent and their consequent while also saying nothing about the nature of that function is to merely restate the problem of conditionals. Gauker says nothing about why some contexts generate functions between antecedents and consequents and others do not. Leaving the existence or non-existence of these functions as brute facts makes it difficult to extend his theory to practical applications. In his examples, Gauker uses contexts whose members are extensionally defined. He explicitly lists all the members of each set. That might suggest that contexts are defined extensionally. But this cannot be so, for it would make his theory patently absurd; language-users are not handed an infinite deck of context-lists by "the world" that explicitly state which elements belong in which contexts when they intersect with specific situations. Clearly, Gauker must be using these extensionally defined contexts for instructive purposes only. His failure to provide demonstration for how one would intensionally define the members of a context severely limits the practical applications of his theory.

One possible interpretation is to say that the world limits the range of values that a given context can take, and it constructs the contexts, assigning elements to each subset. Gauker's most precise definition of how this operates is this:

> The reason why a context qualifies for membership in the multicontext pertinent to a conversation is never just that the interlocutors regard it as belonging. Given the goal of the conversation and the pertinent kind of indeterminacy, it is still the real character of the world in which the conversation takes place that determines which range of contexts best serves that goal relative to that indeterminacy. (Gauker 39)

Indeed, this seems to be the role that "relevance" plays in Gauker's theory, for he repeatedly states that the elements of a context depend upon the goals of the interlocutors (9-11). If so, then it would seem that Gauker's theory fails on its own criteria. In distinguishing his notion of a context, he asserts that pragmatic considerations are unnecessary for evaluating conditionals (Gauker 6).

Instead, Gauker asserts that his inferential, purely lexical contexts are the only type "that matters for purposes of evaluating conditionals" (5). Yet he also states that the world determines the range of values that a context may take. If contexts rely upon real-world *de re* pragmatics for defining context membership, then we have a contradiction. Both would matter for the evaluation of conditionals, since one is used to construct the other. He employs in his analysis the very notion of shared assumptions about the goals of the interlocutors that he had hoped to avoid.

In more recent work, Gauker has eliminated all considerations of speaker intentions and of *de re* pragmatics in the evaluation of language, perhaps to ameliorate problems such as the ones I have suggested. "The proposition that a speaker's words express in context never depends at all on what the speaker intends in speaking" (Gauker 2006, 1). Propositions exist as ordered pairs of context and word. He gives an example of how even a demonstrative doesn't need *de re* pragmatics for interpretation. For example, suppose we are having a conversation, and as a bird flutters by I say

(14) That is a swallow.

This is understood solely by inferential considerations, not by reference to the world. Only in the case of a demonstrative, there is an ordered pairing of context and word. For example, the ordered pair for (14) would be (our context, "that"). It is not to be understood by reference or by actual demonstration (Gauker 2006, 4-7). Nor is it to be understood by my audience recognizing an intention to refer to a bird. (14) is assertible if and only if there is a mapping in our domain of discourse between the ordered pair (our context, "that") and the word "swallow" in the larger domain of discourse, where our context is a subset of the larger domain of discourse. The world, presumably, still defines the domain of discourse. He extends the same type of reasoning to interlocutor goals. "The goals of the interlocutors when a conversation takes place" is itself just shorthand for signaling another inferential rule of language. Again, he does not specify the nature of this inference other than to assert its existence. While this frees his theory of failing on its own criteria, it makes it is even less clear how one could apply it in the analysis of real language. It is not clear how existential claims ("no unicorns exist") could enter contexts on this account of pragmatics.

Goodman's Theory

There is another interpretation of "relevant" that wouldn't cause the theory to fail on its own criteria. Gauker's project revolves around mapping linguistic

items with other linguistic items, not with mapping language. Perhaps a more charitable way to disambiguate what he means by the relevant parts of "the real character of the world" is to interpret it in terms of a minimal set of *ceteris paribus* statements that are necessary to define the antecedent and consequent subsets of the context. Suppose I say to you

(15) If you pour sugar in water, it will dissolve.

On this interpretation, (15) is assertible whenever the antecedent set contains not only "you pour sugar in water," but also a set stating the necessary background conditions for the conditional to be true. For (15) might entail "the water isn't frozen," "the water isn't already saturated with sugar," etc. It isn't just the antecedent alone that must map to the consequent for a conditional to be assertible, but rather the antecedent in conjunction with the set of necessary background conditions that must map to the consequent.

Goodman tried to unpack the meaning of conditionals in terms of *ceteris paribus* causal laws and statements of them (5-30). Goodman argued that a conditional is true when the consequent follows from the antecedent when the antecedent is held in conjunction with a minimal set of sentences that states the relevant background conditions necessary for the evaluation of that conditional (5-30). So (15) would be true only when the antecedent intersects a set of other statements, such as "the water isn't frozen" or "the water isn't already saturated with sugar." The consequent doesn't follow merely from logic alone, but only if certain conditions obtain. Suppose I pull out a match and say

(16) If I strike this match, it will light.

What I mean is that relative to a set of certain conditions, S, being met, the consequent "this match lights" can be inferred from "this match is scratched." So $A \Rightarrow C$ is true only if $(A \bigcap S) \Rightarrow C$ is true. Goodman attempts to flesh out the nature of what the set S might be. But defining the set of *ceteris paribus* conditions is extremely difficult, even for the simplest of conditionals. Goodman concludes that S must be a statement of all the causal laws relevant to that conditional (7-8). But this is far from easy. As Lewis points out, "the principal problem is to specify which further premises X_1, … X_n are suitable to be used with a given antecedent and which are not" (66). S would have to include "the match isn't wet," "the match isn't defective," "there's sufficient friction on the scratching surface," "the match is struck in an oxygenated atmosphere," "the matchstick doesn't break before it lights," etc. S would include not only all the necessary truths, but will also include all the necessary falsehoods. In the end, S would have to include all the necessary statements about the universe. If an

innocuous statement about a match requires that one state all the universe's causal laws, then how is one to analyze a more debatable conditional?

Reduction to Goodman

Gauker's notion of a context suffers from a vicious decision problem about context membership. Are contexts defined extensionally or intensionally? Gauker doesn't say, but it is important to note that in all of his examples, he explicitly lists all of the assertions that are in the context and in the relevant sets. The fact that all of his actual examples employ contexts whose members are extensionally defined suggests that for Gauker contexts must be defined extensionally. But this cannot be so, for it would make his theory patently absurd; language-users are not handed an infinite deck of cards (a domain of infinite discourse?), each of which explicitly states which literals belong in which contexts. Clearly, Gauker must be using these extensionally defined contexts purely for instructive purposes. His failure to provide demonstration for how one would intensionally define the members of a context represents a major practical oversight in his theory.

Suppose I pull out two matches from my pocket, dip one in water without you seeing and leave the other dry. I then set the two matches upon the table. You pick up one of the matches and say

(16) If I strike this match, it will light.

Have you uttered an assertible conditional? Clearly, it seems that it is the world, and not the structure of the discourse, that will make (16) assertible. Whether or not "this match" maps to "lights" depends upon whether or not it is wet, not upon the linguistic structure of our discourse. The linguistic structure of our discourse doesn't signal one particular match over the other when you say "this match." There is nothing on Gauker's account to tell us how these contexts are constructed.

Onomatopoeia presents a further problem for Gauker's semantics. How do words that mimic sounds in the world enter contexts? Suppose we are watching television when our television apparently short-circuits. I decide to unplug it from the outlet since I'm pretty sure that the current is dead. But just in case it isn't, I'll need you to knock me loose if I start to get electrocuted. I tell you:

(17) If you hear "bzzzztk-bzzzztk-bzzzztk," then hit me with the broomstick.

Examples like this suggest that the world constructs the contexts. If so, then while Gauker avoids explicitly using the $(A \bigcap S) \Rightarrow C$ construction, he would

implicitly define membership of his contexts by dint of S, as the real character of the world when an utterance takes place. Gauker uses an analog of Goodman's problematic set, S, of relevant background conditions to construct contexts. The same problem that plagued Goodman's theory resurfaces in Gauker, except transferred up one level. What are the relevant features of the universe that are considered *ceteris paribus* for constructing a specific context? Gauker describes a context as "comprising the essential basic facts pertinent to the world of the interlocutors and the basic features of their situation" (15-26). First, how are "essential basic facts" different from Goodman's relevant background conditions? Second, Gauker defines one vague notion (relevance) in terms of another vague notion (basic facts). The notion of basic facts entails epistemic considerations. Are these basic facts causally related to the consequent or just logically related? If they're logically related, then his account reduces to the truth-functional account. If they're statements of causal laws, then his theory reduces to Goodman.

Fails on Its Own Criteria

Suppose I told you which of the matches was wet. Then the obvious response would be that (16) is assertible just in case I recognize that by uttering "this match" you are referring to the dry one. That would suggest that we define contexts by mutually recognizing each other's intentions in the conversation. But Chisholm has already argued for this in his theory (97-105). Chisholm argued that Goodman's set S can be reduced by recognizing a speaker's intentions in uttering a conditional. The set S could be limited to only those facts that support what I intend to convey. "We can usually tell, from the context of a man's utterance, what the supposition is, and what the other statements are with which he is concerned" (Chisholm 97-105). Although there are problems with this approach to determining the relevant members of a context,[18] it is one that cannot be open to Gauker. If Gauker did respond that contexts are defined by our recognizing what each of us intends to convey, then his theory would reduce to a restatement of Chisholm's. It would also rely upon the very notion he explicitly set out to avoid—shared assumptions about speaker intentions (Gauker 5). Perhaps Gauker might respond that context membership is an analytic primitive, that you cannot reduce analysis beyond it. But that would leave the very problem of conditionals as a primitive of his theory, which is the very problem that he explicitly claims to solve. Restating the problem doesn't solve it.

[18] See Bennett 304-312.

Another problem arises when Gauker states that "a conversation may require spatial and temporal reference points" (256). He describes a scenario about two interlocutors whose conversation relies upon "spatial reference points" to define the context relevant to their conversation. In his example, the two interlocutors designate their spatial reference point through the use of a demonstrative, i.e., through direct extension into the *de re* pragmatics of their conversation (Gauker 256). This would be a fair response for someone who hasn't already claimed that the only necessary contexts for analyzing a conditional are those named by a non-referential domain of discourse (Gauker 5). Gauker's explanation by use of "temporal reference points" seems to violate his own claim that the only relevant concerns are those that belong within his inferential domain of discourse. If non-inferential information coming from the physical world is needed to construct the contexts from the domain, then the physical context of a situation is very important when analyzing a conditional.

Undoubtedly, these concerns are the reason Gauker included his assertion that the goals of the conversation and the real-world features are vital to the analysis of conditionals. The world and the pragmatics of the situation define the domain of discourse. The domain of discourse defines the non-referential contexts and only these contexts are necessary for the evaluation of conditionals. But since features of the real world construct the contexts, saying that real world concerns are irrelevant is a contradiction.

Conclusion

Gauker replaces the notion of logical validity with contextual assertibility, a concept too vaguely defined to do the heavy lifting his theory asks of it. Gauker's theory attempts to describe assertibility in language similar to mathematical logic. Contexts are sets and words and phrases are elements within these sets. Sentences name functions between sets that are members of an inferential domain of discourse. But in order to get English to fit into his theory, Gauker has been forced to make the analogy imprecise: he doesn't state how contexts are defined other than to say that they are defined by the world. The only interpretations that add precision to these concepts reduce it to either an earlier theory or to a restatement of the problem of conditionals. Gauker has a dialectical obligation to other theorists to define these portions of his theory in a non-contradictory manner without causing it to reduce to a restatement of an earlier theory.

Works Cited

Bennett, Jonathan. A Philosophical Guide to Conditionals. Oxford: Clarendon Press, 2003.

Brandom, Robert. Articulating Reasons. Cambridge, MA: Harvard University Press, 2000.

Chisholm, Roderick. "Law Statements and Counterfactual Inference." Analysis 15 (1955): 97-105.

Gauker, Christopher. "Conditionals in Context." Erkenntnis: An International Journal of Analytic Philosophy 27 (1987): 293-321.

———. "Objective Interpretationism." Pacific Philosophical Quarterly 69 (1988): 136-151.

———. Words without Meaning. Cambridge, MA: MIT Press, 2003.

———. Conditionals in Context. Cambridge, MA: MIT Press, 2005.

———. Christopher Gauker's Homepage. 2006. 2006 <http://asweb.artsci.uc.edu/philosophy/gauker>.

Goodman, Nelson. Fact, Fiction, and Forecast. Cambridge, MA: Harvard University Press, 1987.

Grice, Paul. Studies in the Ways of Words. Cambridge, MA: Harvard University Press, 1989.

Jackson, Frank. "On Assertion and Indicative Conditionals." The Philosophical Review 88 (1979): 565-589.

Lewis, David. Counterfactuals. Malden, MA: Blackwell Publishers, 1973.

Lycan, William. Real Conditionals. Oxford: Clarendon Press, 2001.

Interview with Stephen Darwall, University of Michigan

MATTHEW NOAH SMITH
Yale University

Stephen Darwall is the John Dewey Distinguished University Professor of Philosophy at the University of Michigan. [Editor's note: since this interview was conducted, Darwall has been named the Andrew Downey Orrick Professor of Philosophy at Yale University. The University of Michigan has designated him the John Dewey Distinguished University Professor Emeritus.] His research has centered on the foundations and history of ethics and moral theory, and he is the author of several important works in these areas, including: *Impartial Reason* (1983), *The British Moralists and the Internal 'Ought': 1640-1740* (1995), *Philosophical Ethics* (1988), *Welfare and Rational Care* (2002), and *The Second-Person Standpoint: Morality, Respect, and Accountability* (2006).

This interview was conducted for *The Yale Philosophy Review* by Matthew Noah Smith, Assistant Professor of Philosophy at Yale University, whose work focuses on political theory and the philosophy of law.

MATTHEW SMITH: *Describe your early work in ethics and history of ethics.*

STEPHEN DARWALL: I think I have always been interested, from the time I was an undergraduate at Yale if not before, in questions about the "sources of normativity," as Christine Korsgaard calls it: What can make it the case that we ought to do something and, in particular, that we are morally obligated to act? My dissertation as a graduate student at the University of Pittsburgh was about (normative) reasons for acting: What makes some consideration a reason for somebody to do something? And I was always interested in issues about the nature and authority of morality: What kind of reasons for acting do moral obligations purport to provide and do these reasons really exist? I was very lucky in graduate school and my early career to be exposed to a number of philosophers who had thought hard about these issues—Kurt Baier, William Frankena, and W. D. (David) Falk, who was my colleague at the University of North Carolina. My first book, *Impartial Reason*, brings together whatever progress I had been able to make on these questions in the decade after I finished graduate school. One thing I am pretty proud of is that whereas the orthodox view, which I was then criticizing, was that all reasons for acting must

be grounded in the agent's desires, the pendulum has now swung away from that view in the direction of the view for which I argued, namely, that desires themselves are typically based on reasons rather than *vice versa*. I'm not saying for a minute that I played any significant role in turning things around, but at least from our current perspective, it looks like I was on the right side. No doubt things will look different twenty years from now.

I was very fortunate also that the philosophers from whom I learned the most, especially Frankena and Falk, but also John Rawls, whose work influenced me a *lot*, all had a profound sense of the importance of the history of ethics—including for contemporary moral philosophy—and had themselves carefully studied the historical figures who had thought most insightfully on issues about normativity and moral obligation. I also had the good fortune to study with J. B. (Jerry) Schneewind at Pitt when he was just beginning to get interested in the history of seventeenth and eighteenth century moral philosophy. Schneewind would go on to write the best book ever written on the subject, *The Invention of Autonomy*, but when I studied with him he was just getting into it and was known for his work on Sidgwick. In any case, after my first book I became very interested in the thought of the early modern British moralists—some well-known, like Hobbes, Locke, Butler, and Hume, but others reasonably obscure, like Lord Shaftesbury, Francis Hutcheson, Richard Cumberland, and Ralph Cudworth—on the very same fundamental issues of normativity and moral obligation that I had been concerned with in my own work. I found there a treasure trove of philosophical riches, most of it vastly underappreciated, including some fascinating ideas concerning the connection between normative reasons and motivation. My second book, *The British Moralists and the Internal 'Ought'*, came out of this work. Among other things, it argued that the metaethical view we nowadays call "(existence) internalism"—the thesis that something can be a normative reason for someone to act only if it can engage her motivationally (perhaps under certain conditions)—the "internal ought"—is first put forward among these figures and for two very different kinds of reasons. One had to do with empiricist naturalism as a basic philosophical orientation, and the other derived from a view about the relation between practical reason and autonomy. Although Kant is the most well known (and of course best) example of the latter kind of view, it was something of a discovery to find seeds of it in Cudworth, Shaftesbury, and Butler. I think we can see these same two trends represented today, for example, in the work of Bernard Williams and Christine Korsgaard, respectively.

MS: *When I was in graduate school, a lot of people I talked with described you as the philosopher who most appreciated the best in both Hume's and Kant's meta-ethical and ethical theories. Would you describe, in general, how you think Hume's views relate to Kant's views?*

SD: Kant is supposed to have remarked that Hume woke him from his "dogmatic slumbers." I think Kant's Humeanism, if we can call it that, that is, his rejection of any (naïvely) realist metaphysics, holds as much for his practical philosophy as it does for his theoretical philosophy. Kant says somewhere towards the beginning of *The Critique of Practical Reason* that practical reason is fundamentally concerned not with any object that is somehow "given" to the will (say, the good as perhaps a Platonist or a rational intuitionist might hold), but with something deep in the agent herself, as, Kant believes, in the form of the will. This is Kant's Copernican Revolution in practical philosophy that is analogous to his "revolutionary" "critical" theoretical philosophy in *The Critique of Pure Reason*. Hume agrees that central philosophical ideas (causal necessity, for example, in theoretical reason; moral necessity or obligation in practical reason) have their source in the mind. The main difference between them is that whereas Hume takes this mental contribution to be a contingent projection, Kant holds that it is governed by norms of reasoning to which we must be able to conform to be able to reason theoretically or practically at all. I think Kant is right that our idea of moral obligation has the kind of universalizing authoritative ambition he supposes. Moral obligations purport to bind all persons and to give them reasons for acting that are independent of and override their own self-interest or their morally optional ends, values, and projects. And I think Kant's instinct to try to vindicate this ambition by arguing that presuppositions of practical reasoning commit us to it is a correct one. My current view, however, is that he is mistaken to think that what commits us to the universal binding authority of morality are presuppositions of just any first-personal deliberation. As I now see it, it is presuppositions of *interaction*, of engaging one another second-personally and making claims on each other that commits us to this idea. More on that later.

What I think Hume gets right is the role of sympathy in generating benevolent concern and that through sympathy and benevolence we recognize reasons to benefit one another that are entirely independent of any moral obligations we have to one another. What he is not so good on, in my view, is the nature and source of moral obligation and its connection to respect for the dignity of persons. This is Kant's great insight: any person has a dignity and is to be respected as such, regardless of his other merits or even of how he conducts himself as a person—and this fundamental fact underlies morality and moral obligation. I now think this idea is to be interpreted somewhat differently from the way Kant does. As I would put the point: any person has, by virtue of having

the psychic competence necessary to enter into relations of mutual accountability (*second-personal competence*), the standing or authority to make claims and demands and so hold people (himself included) accountable (*second-personal authority*).

MS: *About thirty years ago, you wrote a very famous paper*[19] *making a distinction between two kinds of respect. Would you explain the thesis of this paper? Do you see both kinds of respect in Kant's ethics? Do you see them in Hume's?*

SD: I'll take your word for the "famous" part, but I'm pretty convinced that whatever fame the paper enjoys is an instance of one's name becoming more recognizable through sheer length of time of being on the scene, the passing of an earlier generation, etc., and then people assuming that anyone this well known must have written something of some interest. People tend to fix on some particular thing or things that the newly elders, if I can put it that way, wrote that must justify their recognizability. I always thought the point of that paper was pretty obvious. In fact, I remember Warner Wick, the Editor of *Ethics*, saying as much when he wrote me accepting the paper (though he did admit that no one else had made the point in so many words and accepted the paper for that reason). I also remember his commenting on the paper's "prolixity," as he put it, which could no doubt be said of pretty much everything I write, including these responses. I find that I can't find avoid a certain digressive style, so I've decided to relax and have fun with it. But I digress.

In any event, the fairly obvious point was a solution to a simple puzzle about respect, namely, that we tend to think both that every person is entitled to respect simply by virtue of being a person and that people deserve more or less respect by virtue of how they conduct themselves as persons. Simple solution: there is no conflict because two different kinds of respect are involved. The kind of respect to which we are all entitled is *recognition* of our dignity as persons ("recognition respect," I called it), which we give one another in our deliberations and conduct toward each other by taking proper account of the constraints that the dignity of persons places on our conduct. The kind of respect that we can earn or deserve more or less by virtue of our conduct is a kind of *appraisal* or esteem of our character (so I called it "appraisal respect").

You ask a good question about the role of these different kinds of respect in Hume and Kant, since I think it reveals a deep difference in their views. Appraisal respect is quite important to Hume; it is the appropriate response to

[19] "Two Kinds of Respect," *Ethics* 88 (1977): 36-49. Reprinted in *Ethics and Personality*, ed. John Deigh (Chicago: University of Chicago Press, 1992), pp. 65-78; and in *Dignity, Character, and Self-Regard*, Robin S. Dillon, ed. (New York: Routledge & Kegan Paul, 1994), pp. 181-197.

his central ethical category: virtue or "merit," as he calls it in his second *Enquiry*. In my view, however, Hume is quite tone deaf to recognition respect for the dignity of persons. We can see this in his account of justice. I would argue that a central aspect of justice—indeed, in the very sorts of cases Hume discusses, having to do with property, promises, and contracts—is its connection to *rights* we have that are anchored in the dignity of persons. When we violate these rights, we fail to accord one another a fundamental recognition (respect) that we can claim from each other just by virtue of being persons. For Hume, however, justice and rights have no fundamental standing; they derive from conventions we adopt for mutual advantage. And Hume is quite candid that "were there a species of creatures intermingled with men, which, though rational, were possessed of such inferior strength, both of body and mind, that they were incapable of all resistance, and could never, upon the highest provocation, make us feel the effects of their resentment," though we should give such beings "gentle usage," we would "not, properly speaking, lie under any restraint of justice with regard to them" (Second *Enquiry*). The rough idea is that if someone is not sufficiently strong to resist our abuse of him, then he can make no claims of justice on us. Under these conditions, it is no longer advantageous to restrict our conduct toward him by rules of justice—we can abuse him with impunity (though it would still be vicious of us to do so). There is a lot in Hume that I admire, but this idea I find quite repugnant and simply uncomprehending of the foundation of justice and rights in the dignity of the person. I should emphasize, however, that Hume is an absolutely marvelous philosopher—clearly the best philosopher to have written in English.

Naturally, I think that Kant is better on these points. The idea that any person is to be respected as having a dignity that is "beyond price" is both one Kant clearly articulates, perhaps for the first time in human history, and a central thesis in his moral philosophy. Kant famously claims that a person is an "end in itself" and may never be used simply as a means. I'm not always sure, however, that Kant would himself have interpreted and developed this idea in ways I find most compelling, or even, indeed, in ways that have been identified as prototypically "Kantian" over the last thirty or forty years. When you read the passages on respect and dignity carefully it can sometimes seem as though what Kant has in mind is a precious capacity we have that sets us above the "brutes" that we should look up to rather than a fundamentally equal authority we all have with respect to one another.

MS: *You've recently published a book about what you call the Second Person Standpoint, which is very much a breakthrough in meta-ethics. Please describe the thesis.*

SD: I'm glad you think the ideas may have some promise. To be honest, I do too. This is the first time in my philosophical career that I have had the feeling not just that I was making some progress in getting a hold of some ideas, but that some ideas have gotten a hold of me. It sometimes feels as if I am just channeling something pretty deep that I have somehow tapped into.

One way into the basic ideas is to start where we left the last question. As I see it, the right way to interpret the equal dignity of persons is in terms of an equal authority we all have to make claims on one another (and on ourselves, as representatives of the moral community, by which I mean not any actual community but something like the regulative idea Kant calls the "realm of ends"). Rawls once said that persons are "self-originating sources of valid claims." I think this is just right, and my work attempts to understand, develop, and vindicate this idea.

Where the "second-person standpoint" comes in is through the notion of "address." The only way to make a claim is to *address* it to someone, and when we do so, we are in a second-personal ("I-you," or as Martin Buber put it, "I-thou") relation to someone. This is a distinctive perspective of thought, which differs from a purely observer's (third-person) perspective and from a first-person perspective when we are not directly engaging other persons. (Note that since it always involves I (or we)-you, the second person perspective is also a first-person perspective—it's just that the reverse isn't true: not every first-person perspective is a second-person perspective.)

A main claim of my book is that many central moral concepts—the concepts of moral obligation, rights, moral responsibility, the dignity of persons, and respect for that dignity—are all "second-personal concepts" in the sense that they all involve the idea of an authority to make (address) claims and demands. It is important also that all of these concepts have a conceptual connection to accountability or answerability. It is, I think, a conceptual truth that what we are morally obligated to do is what we are morally accountable or answerable for doing (as I see it, to one another and ourselves as representatives of the moral community). Similarly, it is a conceptual truth that if you have a right to my doing something, then that is something that I am accountable to you for doing. You have a distinctive standing to claim my conduct, or to release me from it if your right is alienable, to bring me to account if I violate your right, to resent the injury, to forgive me if I apologize, and so on.

In addition to making a conceptual argument about the ineliminable second-personal character of these central moral categories, I also make an argument about what we are committed to when we take up a second-person perspective and address or acknowledge any claim whatsoever. (I should stress that I am focusing only on cases where we see ourselves as putting forward or

acknowledging *legitimate* claims, not exerting mere power or force.) I argue that whenever we relate to someone second-personally in this way, we are committed to presupposing that we and she share a fundamental *second-personal competence* (that she has the psychic capacities necessary to hold herself responsible and enter into relations of mutual accountability) *and* that by virtue of that we share a fundamental *second-personal authority* to make claims and demands of one another at all. (In other words, we are committed to seeing one another as "self-originating sources of valid claims.")

In my view, this is the deep point about the equal dignity of persons, and it is a profound point that informs both the content of morality (that we are morally obligated to treat one another in certain ways) *and* its "form" (that we all have standing as representatives of moral community to demand this conduct of ourselves and of one another). I should stress also that I don't think this restricts the content of morality to obligations concerning our treatment just of persons, as though we had no obligations that concern, for example, other animals or the environment. Of course, that may take what my old friend Laurence Thomas calls "a long and unobvious story." (I *think* he got that marvelous phrase from Bernard Williams.)

MS: *You have contrasted your work in "The Second Person Standpoint" with Nagel's work in "The Possibility of Altruism." Nagel's work really set the stage for how moral philosophy is done, and the third-personal standpoint has really been the accepted perspective for ethics. Do you think we might see the second-person standpoint having a similar impact?*

SD: I'm going to beg off answering your question directly and try to answer it indirectly by quoting from myself (from the *Ethics* symposium on my book) about Nagel's insight and the relation between it and what I say about the second-person standpoint. I say something there about how things might have been different if Nagel had carried forward a line of thought (roughly, my line of thought) that I think emerges naturally from something he says about a famous example in *The Possibility of Altruism* (which I ripped off from him, and he from Hume).

"It may be a useful introduction to my way of thinking to compare something that Nagel says in *The Possibility of Altruism* about the kind of example I frequently discuss. I distinguish two different kinds of reason that you might give someone to remove his foot from on top of yours, one flowing from the agent-neutral badness of your being in pain—a reason that anyone might recognize, for example, through sympathy—and the other deriving from a putatively legitimate claim or demand that you have standing to make of him in particular, either as a victim whose right is being violated, or as a representative

of the moral community having the authority to hold others (and oneself) to a moral obligation not to tread on one another's feet.

"Nagel notes that one might make the following point to such an intentional foot treader: 'The essential fact is that you would not only dislike it if someone else treated you in that way; you would resent it.' I could not agree more. As I would analyze things, one would thereby point out to the other that were he to feel resentment, it would be to him as though he had a standing to demand that his foot not be trod upon, and this would commit him to thinking that others have a similar authority to make a similar demand of him also (unless, of course, he could justify to himself the idea that he somehow has an authority that others lack).

"But note how Nagel goes on to analyze the case: 'That is, you would think that your plight gave the other a reason to terminate or modify his contribution to it, and that in failing to do so he was acting contrary to reasons which were plainly available to him.' For me, however, there is an important difference between the idea that the badness of one's plight creates a reason and the thought that one's legitimate claim or demand not to be intentionally caused to have such a plight does. Any consideration of your plight (whether the fact that you are in pain, or that you are being caused humiliating pain, or whatever) would be an agent-neutral consideration having the same normative force for anyone. A claim-based (and so 'second-personal') reason that might present itself through the feeling of resentment, however, would (apparently) apply distinctively to the resentment's object (and, as I analyze it, its implicit addressee), namely, the specific individual(s) who intentionally caused (or might cause) one to have that plight. ('Hey, you can't do that to me.')

"What Nagel must have been thinking is that imaginative resentment when contemplating being badly treated makes one more vividly aware of how bad being a victim of bad treatment is. But so far as that goes, the fact that one would feel resentment and would object, or that one would warrantedly feel resentment and object, plays no role. According to me, however, in feeling resentment one sees oneself as having a valid claim not to be so treated and therefore sees others as having an additional, second-personal reason not so to treat one, one that supplements any reason provided by the badness of any plight resulting from, or in, being so treated. The latter, as Nagel thought, is an agent-neutral reason that anyone has to prevent or alleviate the plight. The former, as I am arguing, is an agent-relative reason that people have not so to treat others themselves. Its second-personal character derives from the fact that it has to do with a legitimate claim or demand, since it is of the nature of claims and demands that they have addressees and implicitly bid for reciprocal recognition of the authority that legitimates them. Their address comes, as I like to put it,

with an RSVP that invites a reciprocating address that realizes mutual respect and reciprocal recognition (hence mutual second-personality). And its agent-relativity is rooted in this second-personal character, since second-personal reasons all concern agents' relations (indeed, their relatings) to one another.

"Now those of us who cut our teeth on *The Possibility of Altruism* will recall that it was this book that reintroduced issues about agent neutrality and agent relativity back into moral philosophy. (I say 'reintroduced' because there had been an earlier flurry over a half-century before involving Moore, Ross, and Broad.) It is, however, precisely because Nagel analyzed the kind of case we are discussing as he did and sought to ground all normative reasons from an 'impersonal,' 'objective,' or 'agent-neutral' point of view that he was led to pose the problem of the justification, or even indeed the coherence, of agent-relative restrictions or deontic constraints in the distinctive terms he did.

"One way of seeing what I am trying to argue, however, is to appreciate that if Nagel had taken his remark about resentment in the way I am suggesting, namely, by noting that resentment involves the idea of warranted claim or demand and, therefore, of a (second-personal) reason not to treat others badly that is additional to the (agent-neutral) badness of the state of being so treated, or of any state resulting from such treatment, he might not have been led to question the justification or coherence of agent-relative restrictions in the way he did, and the entire shape of the resulting debate on this issue might well have been different."

I prefer, however, not to speculate on what influence this line of thought might have as advanced by me.

MS: *Let's talk a bit about this thesis. Do you think that your view requires acceptance of some sort of rationalism about obligation, authority and other moral phenomena of this sort? Or, to put a finer point on it: why do you think that whether A has an obligation to j, or whether A has authority over B must always be explicable in terms of A having reasons to j and B having reasons to accept A as an authority? Isn't it possible that obligation or authority is a primitive relation? That is, couldn't it be the case that we cannot reduce authority or obligation to reasons? Other explanations of the phenomena are surely possible.*

SD: In one sense, the idea that we cannot reduce authority to reasons for acting is something I accept, indeed insist upon and proclaim! Throughout much of recent philosophy, the tendency has been to attempt to explain and vindicate moral obligations in terms of the weight and priority of the reasons for complying with moral obligations—for example, that moral obligations are categorical imperatives that (purportedly) give us reasons for acting that are independent of our desires and interests and that override these. It is a central

claim of mine that this cannot adequately capture the distinctive character of moral obligation because it does not yet bring in obligation's second-personal character, the fact that what we are moral obligated to do is what we are morally accountable to one another (and ourselves) as representatives of the moral community for doing. It is what can be legitimately demanded of us, what we have an authority to demand of one another and ourselves. We simply cannot capture the idea of moral obligation independently of the idea of this authority. But I also hold that this feature is reflected in the character of the reasons that moral obligations provide, namely, that they provide *second-personal reasons* (we could also say, "authority-based reasons") for compliance. So I agree about the fundamental character of second-personal character, but then hold that this fact gets reflected in the kind of reasons we have for acting. (I think, by the way, that I can also explain the categorical character and overriding weight of reasons to act as we are morally obligated, but that's another story.)

Even so, I think there is another sense in which some forms of authority do depend on reasons. The second-personal authority we share as second-personally competent does not; it is fundamental. What grounds it is that it is an inescapable presupposition of the second-person standpoint and the latter's role in practical reason. But any other authority, for example, any differential authority as might be involved in forms of political authority, or in, for example, a military chain of command of an army of a just society that is organized only for purposes of defense, must itself be justified by reasons that derive from this fundamental basic second-personal authority. The reason this is so is that any authority relation is conceptually related to accountability, and no one can intelligibly be held accountable for recognizing and complying with authority unless it is something he can be justifiably expected to accept, that it would be unreasonable for him to reject. This is obviously a stronger condition than any on the existence of normative reasons generally, and what activates it is the connection between authority and accountability.

MS: *It seems that you think that obligations to others not, e.g., to step on their gouty toes depend upon their capacities to respond with Strawsonian reactive attitudes to, e.g., my stepping on their toes. I know you think I could have that obligation even if they couldn't react in that way. Can you explain this?*

SD: Actually, all I argue in my book is that having second-personal competence is a sufficient condition for second-personal authority, not that it is a necessary condition. I think I have an argument that we have obligations to one another (as second-personally competent). I don't have an argument that we have obligations *to* non-second-personally competent beings, but I certainly believe it.

My metaethical view simply lays out what any who holds this normative view is committed to. If I think that I have obligations to other animals, or to human beings who are not persons, in the full-blooded sense that they have rights against me, then the second-personal character of rights and obligations entails that these beings thereby have an authority that they are not able to exercise, and therefore, that someone would have to exercise on their behalf. But this seems to me exactly what we should think if we think we have these obligations. Of course, more minimally, I can think that I have obligations to treat non-human animals and non-second-personally-competent human beings that are not really obligations *to these beings*, without supposing that they have any second-personal authority. I just have to think that among the things we are responsible to one another for is how we treat non-persons also.

MS: *Finally, do you think that epistemological reasons are second-personal? Why or why not?*

SD: Generally, reasons for belief (epistemic reasons) are third-personal; they are considerations which provide evidence for the truth of some proposition, perhaps for someone who is in the epistemic situation in which the agent happens to be. And epistemic authority is third-personal also; I can fully respect it (in the sense of recognize it) without relating *to* someone in any way. If I overhear a stock tip from two brokers who are talking to each other on the train and act on it, I am according them epistemic authority, but not *acknowledging* it to them in any way.

On the other hand, when you and I discuss what to believe on some matter and you assert some proposition, say, that the Detroit Red Wings are going to win the Stanley Cup, something second-personal is going on: you are, as we might say, "bidding" for my belief. You are making a claim on my beliefs, though not on my will. Testimony is an even clearer case, as many philosophers, including Richard Moran, have pointed out. One is presuming an authority to give another person a reason to believe something, asking that she believe it on one's say so. This is clearly second-personal. I do believe, however, that there is a deep difference between second-personal reasons for belief and the kind of second-personal reasons for acting with which I am concerned. The former are defeasible by third personal epistemic authority. If I know you don't know anything about hockey, then this undercuts your authority to give me a second-personal reason to form my hockey beliefs on your say so. But nothing is comparable is true with second-personal reasons for acting. They are, as I say, second-personal all the way down. Another way of putting the point is to say that how you conduct yourself as an epistemic agent can completely undercut your second-personal epistemic authority in discussions about what to believe,

but how you conduct yourself as a moral agent cannot completely undercut your second-personal practical authority. Assume, for example, that there exists a completely evil person in the U.S. prison in Guantanamo. Make him as evil as you like; this would not, in my view, defeat his second-personal authority to claim due process of law or to object to torture. Even an evil person is, after all, a person, and he has therefore the same standing to demand his fundamental rights as you or I.

Interview with Nathan Salmon, University of California, Santa Barbara

LESLIE F. WOLF
Yale University

Nathan Salmon is Professor of Philosophy at the University of California, Santa Barbara, where he has taught since 1984. His research focuses on the philosophy of language and metaphysics, but he has written in many other areas of philosophy, including the philosophy of mind, epistemology, the philosophy of mathematics, and the philosophy of logic. He is perhaps best known for his work on direct reference theory and modality. In addition to numerous papers, Salmon has written several books: *Reference and Essence* (1981, 2005 with new appendices); *Frege's Puzzle* (1986, 1991); *Metaphysics, Mathematics, and Meaning: Philosophical Papers Volume I* (2006); *Content, Cognition, and Communication: Philosophical Papers Volume 2* (2007). Together with Scott Soames, Salmon co-edited *Propositions and Attitudes* in 1988.

This interview was conducted for the *Yale Philosophy Review* by Leslie F. Wolf, a graduate student in the Department of Philosophy at Yale University.

LESLIE WOLF: *When, and how, did you first become interested in philosophy? How did you come to focus on the philosophy of language and metaphysics in graduate school?*

NATHAN SALMON: The earliest philosophical thought I remember—at around age 6—concerned whether God was omnipotent. (I argued not.) After that I considered whether objects of thought had a kind of existence, whether the sense we have of making free choices was illusory, whether the present is more real than the past, what it was in virtue of which mathematics and logic are necessary. I thought about these metaphysical issues well before I heard anyone talk of such things and well before I learned the terms in which philosophers express them.

As a kid, I also had a very good grasp of grammar—the sort involving analysis of sentences into subject, predicate, direct or indirect object, prepositional phrases, gerund phrases, etc. I quickly became more adept at doing this sort of analysis than my grammar teachers were. Not infrequently I corrected their mistakes and raised subtle grammatical perplexities not

mentioned in the textbook, and of which my teachers were completely unaware and had no idea how to answer.

It wasn't until I went to college, though, that I took my first course in logic and learned to think philosophically in a very rigorous way. I transferred to UCLA from a community college in the academic year of 1971-72. That year my philosophical education made a quantum leap. I was extremely fortunate. I took a course from Tyler Burge on Frege, a course from Alonzo Church in the philosophy of mathematics, a course from Keith Donnellan on the later Wittgenstein, a course from Donald Kalish in set theory, a course from David Kaplan on logic, and a course from Saul Kripke in the philosophy of language. I performed sufficiently well in the set-theory course that several non-logic-oriented graduate students asked me to tutor them, both in that course and in its sequel course on meta-mathematics. By my senior year I knew that I would be going to grad school and focusing on the philosophy of logic and language. I retained a strong interest in metaphysics, however. I never took an actual course in metaphysics, not even in grad school. Metaphysics was downplayed until more recently, even despite Kripke's contemporary classic, *Naming and Necessity*. It was only after I studied that marvelous masterpiece that I combined my interests in the philosophy of language and metaphysics. Church, Kaplan, and Kripke were my greatest philosophical influences, three of the finest minds I've known.

LW: *You have argued that merely past objects and merely future objects, as well as merely possible objects, do not exist—but can nevertheless have properties. In particular, you argue that they can have the property of being referred to or being thought of. So, while Socrates does not exist (since he is merely past), we can still refer to him in English by means of the name "Socrates"—or so you claim. Many philosophers, though, insist that an object can have a property only if it exists. How do you respond to these philosophers? If Socrates does not exist, how can I refer to him and think about him?*

NS: I summarized my doctrine with the slogan, '*Predication precedes existence*'. One of the earliest occasions on which I advocated the position in public was at a 1986 conference in Dubrovnik. In my talk I argued specifically that, although Sir Walter Scott no longer exists, he nevertheless presently has the property of having written *Waverley*. During the discussion following my talk, Tim Williamson observed that some philosophers object on the ground that Walter Scott doesn't currently have properties, precisely because he doesn't presently exist. I replied that this retort is really a kind of concession. If Walter Scott lacks the property of having written *Waverley*, it isn't because he didn't write *Waverley* (since he did). Bob Dylan currently lacks the property of having written *Waverley*, because he never wrote *Waverley*. But it is as true of Victor Hugo as it is of Bob

Dylan that he didn't write *Waverley*, and it is equally true of Scott that he *did* write *Waverley*. To admit that Scott wrote *Waverley* while denying him the authorship of *Waverley* merely on the ground that he doesn't now exist, is to concede what really matters about Scott: that he indeed wrote *Waverley*. To insist that he nevertheless lacks the property of having written *Waverley* is like trying to put the toothpaste back into the tube. The property has just been ascribed to him. Disavowing one's commitments is a way to obfuscate, but it isn't a way to avoid those commitments. The position that Scott wrote *Waverley* but lacks the property of having written *Waverley* has all the marks and trappings of an *ad hoc* prejudice rather than a reasoned conclusion.

LW: *Let's pursue this topic, but shift focus slightly. In your paper "Nonexistence," you argue that there are some genuinely vacuous names—i.e., names that refer to nothing, not even to a nonexistent object. For example, you suppose that the name "Nappy" is introduced into English via this stipulation: "Let 'Nappy' refer to the present emperor of France, whoever that might be, if there is one, and to refer to nothing otherwise." You then argue that "Nappy" is a genuinely vacuous name that denotes nothing whatsoever. On your view, what distinguishes a name like "Nappy" from a name like "Socrates" or "Sir Walter Scott"?*

NS: There are several different sorts of proper names that might all be classified as *non-referring* (or *non-designative).* A term might be classified as *non-designative* if there does not exist anything that it designates. The class of non-designative names, in this broad sense, is much more diverse than one might be inclined to think. 'Socrates' is non-designative not in the sense that it doesn't designate, but in the sense that what it designates does not exist. I call this a *weakly non-designative* name. What 'Socrates' presently designates used to exist but doesn't presently exist. Analogously, David Kaplan's 'Newman 1'—a term whose referent is fixed by the description 'the first child to be born in the 22nd century'—designates something that will exist but doesn't. It too is weakly non-designative. I have argued that it is possible for a name to designate a composite object that never exists, though each of the components sometimes exists. I also invented a name, 'Noman 0', which designates something that has never existed and never will, though it might have. I argue that it is even possible to name a composite object that itself could not exist—an impossible object—although each of its components might have existed. The most radically non-designative kind of name is one like 'Nappy', which simply doesn't designate at all, since nothing at all—no possible object and not even an impossible object—is actually, presently emperor of France.

A true singular negative existential is a sentence equivalent to '*a* doesn't exist', where the term '*a*' is non-designative. Some true negative existentials, like

'Socrates doesn't exist', 'Newman 1 doesn't exist', and even 'Noman 0 doesn't exist', are true because the object designated by the subject-term (e.g., Socrates) is something that doesn't exist. This is in sharp contrast to 'Nappy does not exist', a horse of a different color entirely. I argue in a forthcoming paper that this last sentence is true only insofar as it is read in a special way, as expressing that a certain structurally challenged proposition (the "damaged" or "gappy" proposition that ___ exists) is untrue.

LW: *Analyticity continues to be a hot topic in analytic philosophy. In several of your papers, you distinguish between pure semantics and applied semantics, and you define an analytic sentence as a (true) sentence whose truth-value is a logical consequence of pure semantics alone. Can you explain your distinction between pure semantics and applied semantics? What role do you think analyticity has in philosophical methodology and philosophical knowledge? How does linguistic competence relate to analyticity on your view?*

NS: We can say of a proposition that it is true, or that it isn't. It is true that snow is white, for example, and not true that snow is blue. A proposition's being true or not isn't a matter of semantics; it is a matter of metaphysics. We can also say that sentence is, or isn't, true (in a given language). That is to ascribe, or to deny, a semantic property. However, that 'Snow is white' has the semantic property of truth isn't *merely* a matter of semantics. It depends equally on what color snow is, an issue having nothing whatsoever to do with language. What is a matter of pure semantics is the following: that 'Snow is white' is true if and only if snow is white. To infer the left side of this bi-conditional from the right side is to invoke a non-linguistic fact (the color of snow) to establish a linguistic fact (the truth-value of a certain sentence). That 'Snow is white' is true is partly semantic, partly non-semantic. Assuming that definite descriptions are singular terms, the fact that 'the author of *Waverley*' designates whoever uniquely wrote *Waverley* if anyone did is pure semantics; that it designates Walter Scott is partly semantic, partly non-semantic.

Insofar as linguistic competence involves knowledge of the pure semantics of a language, a linguistically competent speaker of a language is in an epistemic position to be able to infer of any analytic sentence in the language that it is true in the language. Whether the competent speaker knows the truth expressed by the analytic sentence is another matter. In rare cases, a competent speaker will know of an analytic sentence that it is true and yet be in no position to know the particular truth that the sentence expresses. An example of this is Kaplan's 'Newman 1 will be born in the 22nd century'.

Historically the notion of analyticity has been extremely important in philosophy. It continues to be important, despite the semantic skepticism of a

recent generation of behaviorists. Descartes, a rationalist, relied heavily on analyticity in his epistemological foundationalism, although the concept had not yet been articulated as such. In the 1970s David Kaplan argued that the '*I am*' in Descartes' '*I think; therefore I am*' is analytic and consequently *a priori*. By contrast, Descartes intended to establish his existence *a posteriori*, on the basis of his experiences. (There is considerable confusion on this point even among historians of philosophy. In a certain sense, for Descartes *all* human knowledge is *a posteriori*—including mathematical knowledge—even though not all of our concepts are derived from experience.) Hume, a nearly complete empiricist and a skeptic, relied on a precursor of analyticity ("relations of ideas") in setting out his epistemological challenge. From Hume right up until the 20th century, empiricists have rested great weight on analyticity to reconcile their epistemology with the evident apriority of logic, mathematics, and fundamental philosophy. Quine was a radical semantic skeptic who rejected analyticity. He attempted an alternative empiricist tack, but in my judgment that tack is no more successful than the logical positivists'. In fact, I find Quine's semantic skepticism a good deal less palatable than the positivists' verificationism. I think these contributions of Quine's were unfortunately seriously retrogressive.

It has generally been assumed that all analytic statements are *a priori*. I argue to the contrary that most of Kripke's examples alleged to be contingent *a priori* statements are in fact, and contrary to Kripke's assessment, analytic *a posteriori*, whereas most of Kaplan's alleged examples of the contingent analytic are in fact neither analytic nor *a priori*. I'm currently working on a paper in which I argue that Quine, in proposing his criterion of ontological commitment, unknowingly implicitly committed himself to analyticity, which he tirelessly opposed. I'm working on another paper in which I rely heavily on analyticity to argue against Kripke's favored response to his famous puzzle about belief.

LW: *What is your view of the role of intuitions in philosophy?*
NS: Intuitions are absolutely vital to philosophy. Philosophical theses are assessed in part by subjecting applications (instances) to intuitive judgment. Philosophical assessment isn't just this, however. I don't advocate the "experimental" approach of canvassing or surveying non-philosophers about particular actual or hypothetical cases. In a very broad range of cases, such a survey is undoubtedly philosophically worthless. In some cases one must also assess other theses related to the target thesis in a similar manner, then balance the results from a meta-perspective. Sometimes one must address the meta-question of whether and to what extent our object-theoretic intuitions might be misled or confused. In a great many cases, maybe in most cases, only the tutored and reflective intuitions of an unbiased but philosophically educated agent are of

any value. But even meta-level investigations are ultimately governed by intuitions of one sort or another, as well as by considerations of overall plausibility.

I think the significance of overall plausibility to philosophy, and to knowledge in general, is typically unduly neglected. When deciding among opposing views, other things being equal, the most plausible view overall is decidedly preferable. I also think that in many hard cases for epistemology (our knowledge of other minds, of the existence of an external world, and the like), mere plausibility will suffice for knowledge in lieu of evidence or proof where the latter is unattainable. Even where other things aren't equal, overall plausibility should count for a great deal. In a very broad range of cases, overall plausibility is much more significant than such traditional pragmatic considerations as theoretical simplicity, ontological economy, making successful predictions, explaining a wide range of phenomena, and the like.

Intuitiveness and plausibility are, of course, closely related. As with intuitions, what is relevant is what is overall plausible to that idealized agent—the tutored, reflective, unbiased, philosophically educated agent. This is nicely illustrated by a famous anecdote. (The anecdote is included in Tom Stoppard's philosophical play *Jumpers*. Elizabeth Anscombe personally assured me that the anecdote is absolutely true.) Wittgenstein asked his students why it was once believed that the Sun goes around the Earth. What made this such a plausible hypothesis? Well after all, the students replied, it *looks* as if the Sun goes around the Earth. In response Wittgenstein asked how it would look if it looked as if the Earth rotated around a stationary Sun.

An idealized agent would have recognized that the way things look doesn't favor either hypothesis over the other. We irrationally tend to see ourselves always as the centerpiece, and to a large degree, as unchanging. Intuitiveness and plausibility must be offset against humanly irrational cognitive tendencies. Heraclitus is often represented as having said that one can't step into the same river twice, because the water is continuously replaced with other water. He didn't actually draw this conclusion, which undoubtedly goes too far. Plato evidently misinterpreted Heraclitus' observation that when one steps into the same river twice one steps into different water. But something like this is true of being in the *same place* twice. When one thinks of oneself as being in the "same place" as before, it is eye-opening to recall that the Earth is continuously hurtling in its orbit around the Sun, the Sun is moving around the galactic center, the Milky Way itself is on the move relative to its neighbors in the Local Group. The universe itself is expanding. Where the Earth goes, we ourselves go along for the ride. The universe is like Heraclitus's flowing river, and we are as single hydrogen or oxygen atoms in that river—even assuming that space is relative. From a very

large perspective we never return to the same location. It is unusual to view things from such global perspectives. Philosophy encourages us to do so. Reliance on intuition and plausibility in philosophy not infrequently requires adopting a very large perspective.

LW: *In your paper "Two Conceptions of Semantics," you describe physicalism and functionalism—two popular theories of the mind—as "philosophically timid doctrines." What is your view of the mind? Do you subscribe to, or are you sympathetic with, some version of dualism?*

NS: In the original manuscript I described those doctrines as "philosophically wimpy." The editor or the referee objected that this sounds unscholarly. Rather than simply omitting the phrase altogether (as I think the editor proposed), I modified the wording, replacing 'wimpy' with 'timid'. I wanted to go on record (even though this wasn't my main topic) observing that a wide range of reductionist and deflationary theses and programs in philosophy—including nominalism, the deflationary theory of truth, physicalism, and a host of reductionisms—although they are very popular, are in my view ultimately based on a kind of prejudice, motivated more by a lack of intellectual vision or imagination than by reason, philosophical courage, and an uncompromising quest for the truth. As an undergraduate I went through a physicalist/nominalist phase. I thought the entire world—including not only physics and chemistry but also psychology, literature, mathematics, humor, emotions, beauty, philosophy, everything—ultimately consisted of nothing but matter, physical forces, fields, energy, and the like. But the indiscernibility of identicals kept messing things up. Descartes' arguments for dualism are but the tip of the iceberg. To insist, despite overwhelming conceptual evidence to the contrary, that there are no numbers, or that a thought is somewhere literally inside the thinker, or that there is no metaphysically significant difference between the propositions that snow is white and that snow is blue, or that the humor of a Woody Allen joke is really just a matter of sound waves and their effect on our brains, is to retreat into a simpleminded fantasy, That fantasy is comforting to some who prefer to keep their heads in the sand over dealing with an enormously rich, complex, and only partially understood reality.

That said, I do think that mental phenomena—as well as institutional facts, aesthetic phenomena, and much else—modally *supervene* on such brute facts as molecular configurations, force fields, and the like. That is, it is metaphysically necessary that if all the brute physical facts are such-and-such, then the mental facts (and the institutional facts, etc.) are thus-and-so. But this is very different

from saying that mental facts conceptually *reduce* to physical facts, let alone that they *are* physical facts.

Let me add here that to me, the most repugnant of philosophical doctrines is the general doctrine that truth (of a proposition) is no big deal. This is something about which I feel passionately. There are numerous variations on the central deflationary theme: truth is unreal, or subjective, illusory, redundant, politically incorrect, insignificant, irrelevant, or not an intellectually worthy pursuit. Frequently adherents of the doctrine misidentify truth with some relatively lame surrogate, typically a pragmatist or socio-political ideal. I once attended a talk by a prominent philosopher who admonished that one should never assert a proposition in an intellectual discussion, even if the proposition is true, if one's doing so might be misused by others to promote some politically incorrect end. I was appalled. That a philosopher would say something as suppressive as that is quite inexcusable. The doctrine that objective truth is unreal or of little value isn't just silly, or sophomoric, or inconsistent (although it is all those things). I believe it is a kind of intellectual disgrace. If the doctrine were correct (that is, paradoxically enough, if it were *true*!), history, mathematics, and even science would have little or no value beyond purely instrumental value. There could hardly be any point *at all* to philosophy—other than the benefits of false advertising. The doctrine is especially insulting to those to whom humanity is intellectually most indebted—extraordinarily insightful thinkers like Archimedes, Galileo, Newton, Darwin, Einstein, Gödel. To this extent, the doctrine soils the cosmic legacy of humankind.

LW: *You defend Millianism, according to which the meaning of a proper name (or other simple singular term) is just its referent. On this view, the names "Phosphorus" and "Hesperus" must have the same meaning, since they have the same referent. However, it seems that these names do not have the same meaning, since John may assent to the sentence "Phosphorus is Phosphorus" while dissenting from the sentence "Hesperus is Phosphorus." What is your response to this objection?*

NS: I dubbed this problem 'Frege's puzzle' in my book of that title. In a nutshell my response to the puzzle comes down to this: our cognitive dispositions toward propositions are tempered by how we take those propositions, and especially by whether we recognize them. If one fails to recognize a proposition when apprehending it in different ways, one might agree to it when taking it one way (under one proposition guise) and yet reject the same proposition when taking it another way. On this conception, it is possible for a single person to harbor conflicting attitudes, rationally and without realizing it, toward what is in fact a single proposition, mistaking it for two independent propositions. Strictly

speaking, the attribution 'Jones believes that Hesperus is the same thing as Phosphorus' states simply that Jones believes the proposition that Hesperus and Phosphorus are the same thing. This will be true if Jones agrees to the proposition when taking it at least one way or another. If Jones agrees to the proposition when it is presented as a trivial truism (as by the sentence 'Venus = Venus'), then the belief attribution is strictly true even if Jones fails to assent sincerely to the sentence 'Hesperus is the same thing as Phosphorus' because he does not agree to the proposition when stated that way.

I believe my response to Frege's puzzle has become the canonical Millian position—although, of course, several Millians do not accept my position and offer rival responses of their own.

LW: *Many philosophers claim that the notions of necessity and possibility are analyzable in terms of possible worlds. According to such an analysis, what is necessary is what holds in all possible worlds, and what is possible is what holds in at least one possible world. In several of your papers, you argue that the notions of necessity and possibility cannot be analyzed in terms of possible worlds. Can you explain your argument briefly? Do you think that modal notions have any analysis at all, or do you think that they are brute?*

NS: I think it is clear that modal notions, such as we express by 'has to', 'mere accident', and so on (as in 'It has to be that the sum of two and three is an odd integer' or 'It is no mere accident that 2 + 3 is odd'), are not analyzable in terms of Leibniz's notion of a *possible world*. Those who, like David Lewis, propose to analyze modality in terms of worlds in Leibniz's sense need to drop the modal qualification 'possible' or 'might have' in their conception of a world, on pain of circularity. They are committed to saying that something is possible if and only if it obtains in at least one world, in at least one way for all things to be, and that something is necessary if and only if it obtains in every way for all things to be whatsoever, *whether or not things might have been that way*. The main problem with this analysis is that a great many ways for things to be are such that things couldn't be that way. For example, one way for all things to be—one "world"—includes my being a credit-card account. Since I'm actually a human being, I couldn't instead have been a credit-card account. (This is essentialism.) It follows that any way for things to be that includes my being a credit-card account is a way things couldn't be.

In fact, the analysis goes in precisely the opposite direction. A *possible world* is a world that *might have* obtained; it is a total scenario, or way for things to be, such that things genuinely *might have* been that way. It then emerges—just as Leibniz held—that something is possible if and only if it obtains (holds) in some *possible* worlds (at least one), necessary if and only if it obtains in each and every

possible world. This is an analysis of a *possible world* in terms of modality, rather than *vice versa*.

Possible worlds don't provide an analysis of modality. I suspect that the basic modal notions themselves are un-analyzable. But the converse analysis of a possible world in terms of modality is far from useless. With it we obtain the result that something is possible in a world *w* if and only if it obtains in some world *w′* that is a possible world in *w*, and something is necessary in a world *w* if and only if it obtains in every world *w′* that is possible in *w*. The notion of one world *w′* being possible in a world *w* is what modal logicians call 'accessibility' between worlds. There is a great deal of conceptual confusion about this in the literature. Even someone as clever as David Lewis was confused about this. He didn't understand what modal-logical accessibility is. (He admitted as much, and concluded that there is no content there to understand. Lewis seriously misunderstood modal notions in general.) Yet the idea is really quite simple: To say that *w′* is *accessible to w* is just to say that *w′* is possible in *w*, no more and no less. The modal-logical accessibility relation is a reflexive binary relation. Most modal metaphysicians, by far, believe that accessibility is in fact an equivalence relation. I've argued to the contrary that it isn't transitive. I've also argued that logic is neutral about whether it is even symmetric. I continue to trust that my position will someday become the conventional wisdom, but at the present time it is excessively unpopular to render the issue even as much as controversial. To most modal metaphysicians, accessibility is an equivalence relation, end of story—without so much as a footnote acknowledging my opposing view. That accessibility is an equivalence relation remains the prevailing view even though the arguments for it are demonstrably fallacious. (Those who criticize cherished doctrine must learn to derive satisfaction in ways other than through reasoned persuasion.)

The identity relationships between possibility and truth in some possible worlds, and between necessity and truth in all possible worlds, enable us to read the logic of modality off of the logic of 'some' and 'all', which we understand through the logic of existential and universal quantification (which, in turn, may be understood in terms of the logic of 'or' and 'and'). In a sense, this is exactly the cornerstone of what is known as modal logic, what I call *the logic of what might have been*.

LW: *Let's talk a bit about the day-to-day of your being a philosopher. My question is, How do you determine which projects to tackle?*

NS: At any given time I typically have several projects at various stages of completion, and I can honestly say that I've never been at a loss for ideas for

possible future projects. There is a reason why I have so many projects going at the same time. I recall being enormously impressed when I saw Michelangelo's statue of David in the Accademia Gallery in Florence. It is exceedingly rare that any human being creates something as exquisitely beautiful. Through his masterpiece the artist still speaks to us. Besides being awed, humbled, moved to tears by the statue itself, I was also equally impressed by the number of the artist's unfinished statues along the corridor leading the way to David. I think there is a potential lesson there for all of us lesser mortals. The physical arrangement of Michelangelo's works in the Accademia is like a message handed down to generations through the centuries from the master himself: "To create *this* [the awe-inspiring masterpiece], one does so by repeatedly doing *that* [the unfinished statues]." To achieve the best one is capable of, it can be important to set a project aside and to concentrate one's efforts on a different project entirely—perhaps even one after another after another. After a time, one can return to the original project with a fresh eye and improve upon what one has already done. I'm no Michelangelo, of course, but I endeavor in my work to follow his implicit advice.

LW: *It is clear from your books and papers that you are greatly interested in literature and music. And you play the guitar. Has your interest in the arts influenced your approach to philosophy in any way?*

NS: I'm much more into music than literature. (Although I work on the philosophy of fictional objects, I seldom read fiction. In fact, I've never much liked reading in general. I force myself to read.) I'm a self-taught guitarist, and I sometimes play semi-professionally. I play entirely by ear. I learned to play by listening to the Beatles' recordings which I had (and still have) on vinyl records, and figuring out the chords. I was very fortunate in two ways when I began teaching myself to play. First, I inherited a very good musical ear from my grandfather, who was a superb self-taught musician (unlike me). Second, the gifted pianist/composer, James Newton Howard, was my classmate and friend, and lived only a block or so from me. Although he was a child prodigy, as far as my classmates and I were concerned, "Jimmy" was a regular guy whom we hung out with—when he wasn't practicing piano. He downplayed his musical talent, and although his friends were aware of that talent we set it aside as irrelevant. But when it came to music, Jim seemed to know everything there is to know. He taught me most of what I know about music theory. In 1966 I played a new Beatles album for him—the American-released *Yesterday …and Today* (culled from the British version of *Revolver* and earlier recordings). Although he was by that time already a very accomplished classical pianist, he responded with

obvious enthusiasm for the melodies, harmonies, chord structure, and creative skill displayed on that album. Jim was uncanny in his ability to analyze the songs on the spot. He was in fact a great inspiration to me. To this day I enjoy figuring out a song's chords entirely by ear, especially when those chords aren't obvious. But I don't play nearly enough to be genuinely accomplished.

I keep my interest in music mostly separate from my interest in intellectual matters. I treat music more as a way to clear my head when I become excessively analytical, or when the demands of the profession occasionally become excessively unpleasant. Music is a refuge. Whenever I watch a movie I listen to the score. (I rarely listen to lyrics.) I love the music of Bach, especially when things in my life are in disarray. I think this is because Bach combines great beauty with orderly, analytical precision. But I'm equally fond of the music of Puccini, probably for its combination of great beauty with unrestrained, go-for-broke passion. I'm fond of the music of Ennio Morricone for the same reason.

SPECIAL THANKS TO:
Michael Della Rocca and the Yale Philosophy Department
Seyla Benhabib and the Yale EP&E Department
The Yale Undergraduate Organizations Funding Committee
Andrew Mangino and the *Yale Daily News*
TYCO Copy Center & Fine Stationers
Jordan Corwin and Yaron Luk-Zilberman

This issue of THE YALE PHILOSOPHY REVIEW was produced using Microsoft Word 2008 and was printed by TYCO Copy Center & Fine Stationers of New Haven, Connecticut. The typeface is Garamond.

printed by
TYCO