THE JOURNAL OF
PHILOSOPHY

VOLUME CI, NUMBER 4
APRIL 2004

167 *The Irrationality of Unhappiness and the Paradox of Despair* Sarah Buss

COMMENTS AND CRITICISM
197 *On Floyd and Putnam on Wittgenstein on Gödel* Timothy Bays

BOOK REVIEWS
211 Candace Vogler: *Reasonably Vicious* Henry S. Richardson

216 NEW BOOKS
219 NEW BOOKS: PAPERBACK REPRINTS

***Published by* The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Editor Emeritus*: Sidney Morgenbesser. *Consulting Editors*: David Albert, John Collins, James T. Higginbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor*: John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Akeel Bilgrami; *Secretary*, Daniel Shapiro; *Treasurer*, Barbara Gimbel; *Other Trustees*: Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.
You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2004

SUBSCRIPTIONS (12 issues)

Individuals	$35.00
Libraries and Institutions	$75.00
Students, retired/unemployed philosophers	$20.00
Postage outside the U.S.	$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, *if* the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.
POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2004 by the Journal of Philosophy, Inc. ISSN 0022–362X

THE JOURNAL OF PHILOSOPHY

VOLUME CI, NO. 4, APRIL 2004

THE IRRATIONALITY OF UNHAPPINESS AND THE PARADOX OF DESPAIR*

Happiness is an end in itself. This is so obvious that even those who argue against the single-minded pursuit of happiness take it for granted that, all else being equal, we have good reason to refrain from behaving in ways that make us unhappy: we have good reason to avoid unhappiness because unhappiness is bad.

Even if, however, by its very nature, unhappiness just is the sort of thing that (all else being equal) any rational person would rather take in as small a dose as possible, one might still wonder whether something more can be said to illuminate this fact. In the pages that follow I pursue this possibility. Our natural desire to avoid unhappiness has the support of our reason, I argue, not only because unhappiness is an unpleasant psychological state, nor, more generally, because being unhappy is the opposite of "doing well." Our desire to avoid unhappiness is also justified by a *formal* principle of rationality; for unhappiness is itself a form of irrationality.

The relevant rational principle not only supports our preference for happiness over unhappiness; it also explains why we have good reason to care about our happiness in the way that we characteristically do. We value our own happiness not simply because it is happiness, but because it is *ours*.[1] Other people's happiness matters too, of course.

---

* I would like to thank Barbara Herman, Elijah Millgram, Martha Nussbaum, Connie Rosati, Angela Smith, and David Velleman for their comments on earlier drafts. I am especially indebted to Maggie Little for her patient and probing suggestions. The footnotes do not begin to indicate how much this article owes to her help.

[1] In this article I am thus not interested in what Stephen Darwall calls "self-concern": "self-concern…is only incidentally egocentric—its object being, unlike that of attitudes *de se*, the individual one is, rather than oneself as such…self-concern [is] an instance, in one's own case, of an attitude one can have in principle toward any individual, a concern that the moral point of view is thought to express equally toward everyone"—"Self-interest and Self-concern," *Social Philosophy and Policy*, XIV, 1 (Winter 1997): 158–78, here p. 160.

0022-362X/04/0104/167–96

But if we are honest, we have to admit that, to us, our own happiness (and that of our loved ones) matters more. The principle to which I call attention here sheds light on why we are rational to privilege ourselves in this way. It does not require us to regard our motives as universalizable, nor our selves as one among many. It implies that there are circumstances in which, independent of any other-regarding considerations whatsoever, we must, if we are to avoid irrationality, either engage in self-interested behavior aimed at reducing our unhappiness or revise our beliefs about who we are.

In making my case, I will describe a form of unhappiness—unhappiness as despair—and I will argue that unhappiness in this form has a paradoxical structure that explains why we have reason to try to avoid it. If, as I will also argue, despair is not only an important form of unhappiness, but a pervasive form as well, then the principle of rationality that is hostile to despair implies, more generally, that we ought to do what we can to avoid unhappiness.

At the heart of my account is an aspect of our self-concern that does not get as much attention as it deserves: our "personal ideals." The important role played in our lives by our personal ideals reflects the fact that they combine into one condition-to-be-realized our views about what is valuable and our views about who we are. I will devote much of this article to exploring the nature of our ideals and the implications of the way they integrate our values and aspirations with our self-conception. This exploration will uncover the deep connection between our ideals and our hopes. And this discovery will lead, in turn, to a discussion of what I call "the paradox of despair." To see why despair is paradoxical is to see why our reason opposes it. And with this insight, we will finally arrive at the supplementary justification of our self-regarding desire to avoid unhappiness.

Since we can cease being unhappy without becoming happy, reason's demand to avoid unhappiness is not, strictly speaking, a demand to pursue happiness as an end in itself. But it comes very close. After all, pursuing happiness is the natural way that we human beings flee unhappiness. So becoming happy is a means to becoming less unhappy in the same way, and to the same extent, that becoming sated is a means to becoming less hungry: just as being free from hunger and being sated would be the same thing if it were not for the fact that one could put an end to hunger by killing oneself, so too, being free from unhappiness and being happy would be the same thing if it were not for the fact that one could put an end to unhappiness by killing oneself—or killing all hope.

I. UNHAPPINESS AS THE FAILURE TO REALIZE PERSONAL IDEALS

Happiness and unhappiness mean many different things to different people. What interests me in this article is the familiar, widely shared conception of the experience of (dis)satisfaction, or (dis)contentment, with oneself and one's life. According to this conception, happiness and unhappiness are moods or emotions (not necessarily intense) directed at one's own condition. They are emotional states that pervade one's "being in the world."[2] When one is truly unhappy, a shadow—sometimes dark, sometimes faint—is cast over everything one encounters.

Happiness and unhappiness are not mere qualitative states. Each has a cognitive, as well as an affective, component. We are (un)happy *about* (or *with*) our condition; and (at least in most cases) we *believe* that our situation warrants the feeling of (dis)satisfaction it generates. A person can be (un)happy without perceiving the source of her (un)happiness. But if she is happy, she must believe that "things are good." And if she is unhappy, she must believe that, somehow, "something is wrong."

We are naturally disposed to care about whether we experience our own lives in this way.[3] Though we may rarely think about whether we are happy, and though happiness is rarely the express object of our actions, a human being who is indifferent to her own happiness is a specimen as strange as she is rare. This fundamental mode of self-interest is distinct from our more general concern to live good lives. Our poor health is especially intolerable to us when it "gets us down"; and unhappiness can be provoked by events that do not really prevent us from living perfectly healthy, fulfilling lives.

The unhappy person's judgment that there is "something wrong" with her life is often difficult, if not impossible, for her to flesh out with a more determinate description. Nonetheless, whatever the details, unhappiness is distinct from many other forms of dissatisfaction. When someone is unhappy, it is not simply that some of her desires are unsatisfied—or even that she would prefer it if things were different in certain respects. (Surely, no human being with a working mind could fail to satisfy *these* conditions!) Nor is unhappiness essentially a matter of wanting something *very intensely*—the fancy car in the local lot, a roof over one's head. Rather, the unhappy person's judgment that "something is wrong" is the judgment that her life falls short of what counts as a good life for her—that she is not living the life her personal ideals mark out as "to-be-realized" by her. A person is happy, in the relevant sense, if and only if she has a positive feeling about the gap between her reality as she perceives it and her personal ideals. She is unhappy if and only if she has a negative feeling about the perceived gap. It may, in principle, be possible for someone to believe that no gap exists, and yet feel nothing in response. So, too, someone may be unmoved by even a very large perceived gap. But while these are important conceptual possibilities, they are rare human conditions. In almost every human being, the cognitive component of unhappiness naturally gives rise to the affective component.

---

[2] Clearly, then, the psychological state that interests me in this paper is not the purely formal conception of a final end which Aristotle seems to have in mind when he announces that all human beings want to be happy.

[3] Even if it is better to be Socrates dissatisfied than a pig satisfied, to be Socrates satisfied is best of all—or so we (and Socrates) think.

Because there is no principled limit to the content of our personal ideals, human beings can be unhappy about anything and everything—just as they can be frightened by anything and everything, and just as they can find anything and everything funny, or depressing, or boring. But whatever the particular object of one's unhappiness may be, it makes one unhappy insofar as, and only insofar as, one takes it to fall short of one's current personal ideals. To be unhappy is to feel that things are not, in some respect anyway—perhaps in a way one cannot yet identify—as they must be if one's life is to be a truly good life.

Some human beings are more susceptible to unhappiness than others because they are less tolerant of perceived discrepancies between the ideal and the real, or because they perceive discrepancies as being greater than they really are. The slightest discrepancy evokes misery in some; others are not so easily bothered—whether because they do not notice the discrepancy (or its extent), because they are more optimistic about the prospects of improvement, or simply because they are willing to settle for less. In any case, our happiness is linked to our personal ideals. So, if we are to understand our interest in our own happiness—and if, in particular, we are to understand what distinguishes unhappiness from the standard human condition of living with unsatisfied preferences—then we had better take a closer look at the role these ideals play in our lives.

Most of us take an approving attitude toward many things. There are many things, people, and deeds that we like, admire, judge good. Of these, however, we identify very few as part of our own good. This is because very few strike us as having anything very important to do with us; and this, in turn, is for two compatible and overlapping reasons: (1) they do not matter that much to us, and (2) they are not related closely enough to who we think we are.

We may be overwhelmed with admiration for a person's knowledge of botany, or for a master painting in the Louvre, or for a daring trek to the top of Mt. Everest, without having the least interest in acquiring or doing these good things ourselves. We may value the talents of Michael Jordan without having the least desire to "be like Mike." The point is not that we do not believe that our lives would be better if they contained these good things. To be sure, we might refuse the offer to climb Mt. Everest, even with all expenses paid. But if a Genie asked us whether we would like to know more about plants, or whether we would like to be able to dunk a basketball, we might not hesitate to accept her offer. The point is that, however wonderful we think it would be to have these wonderful things, we do not regard them as to-be-had *by us.* Like the ability to fly, they are very nice, indeed. But they are not essential—or even important—components of our own good. They are not essential to our flourishing. Our lives do not "fall short" because they lack these good things.[4]

The importance of distinguishing between the judgment that something is good and the judgment that something is part of *one's own* good is implicitly acknowledged in the many philosophical attempts to sketch "the good life" for human beings: not every good thing is essential, or even important, to a good human life. Having acknowledged as much, however, we need to go further. For there is more to a good human life than its relation to the human nature we all share; though some human goods are surely universal, not all of them are. Human beings come in many varieties. They differ greatly from one another with respect to their values and their self-conceptions. And so, they differ greatly with respect to the sort of lives that are good lives-for-them, and the sort of lives they take to be good lives-for-them. In short, human beings have different personal ideals; and so, they differ with respect to the conditions under which it is possible for them to be happy. The point is not that no two human beings could possibly have the same personal ideals: our personal ideals are "personal" not in the sense that they are unique to us, but in the sense that they represent *our* conception of the good life *for us.*

As we grow into self-consciousness, who/what we take ourselves to be and what changes we think it is reasonable to expect in our lives are mutually determining. And each, in turn, contributes to, and is constituted by, our personal ideals. From a very young age, we each begin constructing our own (rather fuzzy) picture of who we are and what we are capable of doing and becoming. This involves figuring out not only what it is to be human, but what it is to be *me—this particular human* who differs from all others. The process takes place, moreover, even as we are learning what we like, what we do not like, and what is *worth* liking. Indeed, these two developments are mutually reinforcing: our evolving self-conception influences our conception of what is valuable, and our evolving sense of what is valuable influences our self-conception. The whole complex process generates a conception of our good in the form of a set of personal ideals, or conditions to-be-realized by us. Thus, for example, a young child who comes to think of herself as a "jock" also comes to value being able to run fast and throw far, and to expect that she will be able to run fast and throw far. At the same time, because these accomplishments are among her personal ideals, they contribute to how she thinks of herself; she regards herself as the sort of person who has these ideals; that is, she thinks of herself as a "jock."

---

[4] This is a frequently overlooked reason for rejecting the conception of happiness as the satisfaction of the greatest number of one's desires.

Many of the nice conditions that are not part of our good are conditions we could satisfy if we really wanted to; and many more are conditions we could at least come closer to satisfying. Other conditions, however, are not part of our good precisely because we do not think we can possibly satisfy them. More carefully, *we* do not *take* them to be part of our good because we do not believe that satisfying them is a live possibility for us; our satisfying them is not, we think, something we can reasonably expect to happen.[5] Goods that fall into this category are not conditions to-be-satisfied by us; and so, though we may fantasize about what it would be like to satisfy them, we cannot regard them as goals, or ends, to pursue. Surely, they are components of *someone's* good life; but this good life could not possibly be our own.[6]

As the examples show, many conditions which fail to qualify as part of our good for this second reason can also fail to qualify for the first reason. Indeed, we may come to regard them as not very important precisely because we have no expectation of satisfying them. (This may be the story behind many adjustments to handicaps.[7]) Some disqualified goods, however, may be conditions we think it would be well worth pursuing if there were any reason to expect that this would make a difference. The only reason we do not regard them as part of our good is because we do not believe they could possibly be part of our lives.

---

[5] If there is a *teleological* constraint on our good, then this is it: flying is not part of our good because we are not "made to fly."

[6] "Who," Pascal asks, "is unhappy at having only one mouth? And who is not unhappy at having only one eye? Probably no men ever ventured to mourn at not having three eyes. But anyone is inconsolable at having none"—*Pensees*, W.F. Trotter, trans., in Pascal's *Pensees and The Provincial Letters* (New York: Random House, 1941), #409, p. 130. The possibility at issue here refers to what there is any likelihood that we, or anyone else, could do. It does not refer to what is *conceptually* possible. For more on this point, see pp. 184–86 and note 24 of this article.

[7] For one interesting relevant discussion, see Edward Dolnick, "Deafness as Culture," *The Atlantic Monthly* (September 1993): 37–51. According to many in the deaf community, Dolnick reports, "deafness is not a disability. Instead, many deaf people now proclaim they are a subculture like any other. They are simply a linguistic

As we will see, this does not mean that *alterations* in our conception of what is possible automatically lead to *alterations* in our conception of our good. (If our conception of our own good altered every time we discovered that we were unrealistically optimistic in our appraisal of our possibilities, then we would have no immediate knowledge of what unhappiness is.) What is more, our animal nature imposes limits on the extent to which our conception of our good can reflect our appraisal of our possibilities. (As I will have occasion to stress later,[8] if we found ourselves in circumstances that made it unrealistic to expect to satisfy our most basic needs, most of us would be incapable of concluding that these needs are really not so basic after all, that we do not have to satisfy them in order to thrive.) These qualifications notwithstanding, however, the connection between our perceived possibilities, our self-conception, and our conception of our good goes very deep. Whenever we think that our good is at odds with our possibilities, we experience a psychic pressure to find a new way of

---

minority (speaking American Sign Language) and are no more in need of a cure for their condition than are Haitians or Hispanics.... So strong is the feeling of cultural solidarity that many deaf parents cheer on discovering that their baby is deaf" (pp. 37–38).

For another interesting discussion of the human tendency to adjust one's sense of normalcy (one's self-conception), see Wilfrid Sheed, *In Love with Daylight: A Memoir of Recovery* (New York: Simon and Schuster, 1995), pp. 9–54. "Polio," Sheed says, "caused me to lose something quite irreplacable, something I would have sworn I couldn't live without" (p. 13). But despite having feared that "I'd go crazy if I got polio and had to give up baseball, and of course football—and *walking*" (p. 26), he discovered that "the speed of mental adjustment can be quite uncanny" (p. 19)—that "you get used to things incredibly quickly, and are ransacking the horizon for new pleasures almost before the old ones are out the door" (p. 46). "If," Sheed confesses, "you'd told me just six months before that the height of happiness would soon consist of leaning on a pair of crutches and inhaling deeply, I'd have woken up screaming" (p. 34). But "once you are used to your new condition, your imagination becomes free once again to rest or amuse itself; you stop scanning the skies for miracles, and life returns to an agreeably small scale of operations" (p. 49). Like other polio victims he met, Sheed never thought of himself as "handicapped."

Recently, both economists and psychologists have been very interested in the capacity human beings have to "adapt" to conditions they thought would be incompatible with their happiness. As economist George Loewenstein explains, "when any event occurs to us, we make it ordinary." In other words, we naturally adjust our conception of what counts as ordinary-for-us. See Jon Gertner, "The Futile Pursuit of Happiness," *The New York Times Magazine* (September 7, 2003): 44–47, 86, 90, especially p. 47.

[8] See pp. 195–96.

flourishing.[9] This, I will argue, is the pressure of our commitment to being noninstrumentally rational. It takes the form of our natural impulse to pursue our own happiness.

II. UNHAPPINESS AS DESPAIR

If most things are not part of a person's good, and if, nonetheless, it is possible to be unhappy about anything, then there is one obvious sense in which unhappiness can be irrational: someone's unhappiness is irrational if her personal ideals are irrational. This sort of irrationality occurs when someone has a false conception of what counts as "her life," or when she is confused about what counts as a candidate human good. The borderlines are fuzzy, of course, but someone can, I think, plausibly be deemed irrational if she is made unhappy by the recognition that somewhere, someone has failed an exam, or if she is unhappy that the sky is blue, or that she is in excellent health—where these states of affairs have no connection to anything else that could justifiably make her unhappy.

To the extent that someone's unhappiness thus depends on perverse ideals, she has an obvious reason to do something that is likely to make her happier, namely, change her ideals. But this reason applies only to perverse cases of unhappiness. And more importantly, it is not a reason to be concerned about the fact that one is unhappy, or to be moved by an interest in becoming less so. Accordingly, it does not shed light on the legitimacy of our natural self-concern.

The key to this supplementary justification is, I will argue, the structure that all cases of unhappiness have in common: the unhappy person is unhappy that the reality of her situation falls short of her personal ideals. When she believes, in addition, that this gap cannot be closed any further, then her unhappiness has the form of despair. I will soon argue that contrary to conventional wisdom, this form is the garden variety of unhappiness. For now, it suffices to point out that life affords countless examples—from the young ballplayer who realizes that she cannot make the team, to the graduate student who realizes that she cannot "make it" in her chosen profession; from the woman who finally admits to herself that she will never be able to have a child, to the woman who knows that nothing can bring her dead child back to life. The emotional states of such people can vary widely in intensity. Indeed, a person can be in despair without even being aware of this fact. I want to argue, however, that in all cases, unhappiness as despair is a form of irrationality. A human being in despair is not necessarily *criticizably* irrational, but she is irrational nonetheless.

---

[9] As I indicate in note 29, the inner conflict referred to here can be characterized in Stoic terms: one wants to conform one's aspirations to nature's design of the world, but to do so would be, it seems, to flout nature's design of oneself.

In order to understand this rather dark pronouncement, it is necessary to take a close look at the relation between despair and hope. In particular, it is necessary to consider the paradoxical nature of this relation. To this end, I want to call the reader's attention to Kant's discussion of regulative ideals. According to Kant, regulative ideals reflect the demands of reason. And unless we recognize this fact, we will find ourselves in the grip of a paradox: the hope which these ideals necessarily presuppose will itself give rise to despair.

Kant makes this point in discussing both science and morality. When we do science, we are trying to make sense of natural phenomena. But this means that we must assume that the world is intelligible to us. This, in turn, means that we must assume that nature is a systematic unity. And so, Kant concludes, we must believe that it is reasonable to hope that we will discover this unity, and the unconditioned condition that underlies it.[10] Similarly, when we think about how we have reason to act, we necessarily presuppose that it is possible for us to achieve the only unconditional end: the Highest Good, happiness in proportion to virtue. If this end were not achievable, reason would not demand that we pursue it. So, insofar as we are committed to complying with reason's demands, we necessarily believe that it is reasonable for us to hope that the Highest Good will be realized.[11]

There seems to be a problem, however. For when we take stock of the evidence available to us, neither the scientific nor the moral hope seems to be justified: neither ideal seems to be realizable. No evidence could possibly support the claim that nature is rational; and surely it is far more probable that the natural world is *not* in perfect harmony with our own faculties. Things are at least as bad in the moral context. For there seems to be ample—indeed, overwhelming—evidence that happiness is not apportioned to virtue.[12] It thus seems that we have good reason to despair of ever realizing our ideals.

---

[10] "The law of reason which requires us to seek for this unity is a necessary law, since without it we should have no reason at all, and without reason no coherent employment of the understanding, and in the absence of this no sufficient criterion of empirical truth. In order, therefore, to secure an empirical criterion we have no option save to presuppose the systematic unity of nature as objectively valid and necessary"—Kant, *Critique of Pure Reason*, Norman Kemp Smith, trans. (New York: St. Martin's, 1965), A651/B679, p. 538.

[11] "It is in the view of reason, in the field of its theoretical employment, no less necessary to assume that everyone has ground to hope for happiness in the measure in which he has rendered himself by his conduct worthy of it..."—*Critique of Pure Reason*, A809/B837, p. 638.

In discussing the Highest Good, Paul Guyer calls attention to this basic feature which, according to Kant, every human ideal must have if it is to play a role in our choices: "Kant's position seems to be that while assurance that the noncontradictoriness of an end would be all that is needed from a theoretical point of view to make a course of action aimed at that end rational, human psychology is such that in fact it needs a greater incentive, a positive reason to believe its end is realizable"—"From a Practical Point of View: Kant's Conception of a Postulate of Pure Practical Reason," in *Kant on Freedom, Law, and Happiness* (New York: Cambridge, 2000), pp. 333–71, here p. 364.

Yet if they really are our ideals, then we cannot but hope to realize them. So we seem to be stuck.[13] Again, it is only because we have certain ideals that we can despair of failing to realize them. So a necessary condition for the possibility of their being our ideals is a necessary condition for the possibility of our despair: our despair depends on our hope.

It might seem that one could escape the paradox by simply abandoning the ideals. But this, Kant argues, is not a genuine possibility. And once we see why this is so, we see how to resolve the paradox. It is not possible to abandon our regulative ideals because they are the product of reason itself; without them, there can be no scientific reasoning, and no rational action. This means that hope is a necessary condition for the possibility of reasoning about why things are as they are, and about what is worth doing. Since reason itself demands that we hope—since we cannot reason without hope—we are justified in hoping. In relation to this justification, the lack of supporting evidence is beside the point.

Kant's discussion of regulative ideals is, of course, far more nuanced than this crude sketch suggests. But for my purposes, it is enough to point out the light this discussion sheds on the nature of the ideals that especially interest me in this article—and on their relationship to hope and despair. Kant shows us that experiencing despair over the impossibility of realizing the regulative ideals of reason presupposes the hope that we will, indeed, realize them. Similarly, I believe, when we are unhappy at the thought that one of our personal ideals cannot be realized, this is precisely because we still hope to realize (or at least more nearly realize) it. Our despair at the unrealizability of our personal ideals presupposes that they "regulate" our choice of ends; and their playing this role presupposes the hope that we will, indeed, realize (or at least more nearly realize) them.[14]

---

[12] This evidence also appears to count against the assumption that the ends of nature form a systematic unity.

[13] Allen Wood puts the situation very nicely: "Just as the dialectic of pure theoretical reason produces an illusion which 'unceasingly mocks and torments' us in our pursuit of knowledge, so the pursuit of the ideal of pure practical reason will lead us necessarily into the troubled waters of illusion also, where we will be threatened with the unattainability of an ideal which we ourselves cannot establish, but with which we cannot cease to concern ourselves without forsaking the rationality which is proper to our own nature"—*Kant's Moral Religion* (Ithaca, NY: Cornell, 1970), p. 98.

The best way to underscore the Kantian lesson that despair requires hope is, I think, to consider cases. Take Jane, for example, who has spent many years studying classical ballet. She thinks of herself as a dancer, and in so doing, she adopts the personal ideal of being able to do all the things that a dancer must do. Having adopted this ideal—having come to regard dancing as to-be-realized by her—she reinforces her conception of herself as a dancer. And in adopting this self-conception, she thinks of herself as having the capacities of a dancer: if she works hard enough, she believes, she will be able to twirl and leap gracefully, and do all the other things a dancer must do.

She believes this. She really does. But lately, things have not been going so well. Jane refuses to deceive herself. "My dancing," she thinks, "is lousy. It falls short of what it must be if I am to be a dancer." This fact challenges the self-conception implicit in her personal ideal: it challenges her belief that she is a dancer, that dancing is a part of her good, that in dancing she realizes her potential—in short, that she can (with hard practice) do the things that a dancer must do.

Not only is Jane unhappy. She is at a crisis point in her life. It might seem that the obvious way for her to resolve the crisis is to give up the ideal of being a ballet dancer. But it is important not to underestimate the difficulty of doing so. Giving up a personal ideal involves renouncing something one values highly. Worse still, it involves abandoning the self with which one has identified in adopting this ideal. This can be extremely difficult—even impossible. After all, our self-conceptions and the identifications they imply are not wholly responsive to our will; we cannot simply *decide* to be someone else once we realize that we are mistaken about who we really are.[15] Not only, moreover, are such shifts in self-conception psychologically difficult; they are also in tension with the virtues of loyalty, faithfulness, and constancy, and of courageously "sticking up for what one believes in." A person thus has good (moral) reasons, as well as strong motives, for holding on to her ideals, and to the hopes they support.

---

[14] As Harry Frankfurt explains, a person "*guides* himself by reference to [his ideals]"—"On the Necessity of Ideals," in *Necessity, Volition, and Love* (New York: Cambridge, 1999), pp. 108–16, here p. 111. My point here is closely related to a familiar point about desires. As Alfred Mele explains, "If I am convinced that I cannot travel faster than the speed of light, or change the past, or defeat the current heavyweight champion of the world in a fair fight, then although I might wish that I could do these things, I do not *desire to do* them. Achieving the represented objects of action-desires is doxastically open for the agent: if she does not explicitly believe that she can A, at least she is not convinced that she cannot A. Any desire to A ('A' being an action variable), by its very nature, inclines the agent, in some measure, to A intentionally, or to try to A, or to try to put herself in a position to A. This is part of what it is to be an action-desire"—"Motivational Strength," *Noûs*, XXXII, 1 (1998): 22–36, here 25. My point is that not only are our "action-desires" sensitive to our beliefs about what is possible, but that insofar as our ideals are not mere wishes, but actually *guide* our action-desires, they are under pressure to conform to our expectations. It is not (as I will soon have occasion to stress) that they cannot possibly defy our expectations, but that insofar as they do so, we have an incoherent/conflicted self-conception.

Understandable, and even justifiable, though Jane's hope may be, the evidence clearly indicates that it is groundless. Because Jane knows this, she despairs. She despairs of ever being the dancer she still hopes she might be. She despairs because she still believes that being a dancer is part of her good, because she still has the self-conception which is linked to this conception of her good, because she still accepts the possibilities implicit in this self-conception, and because she clings to these possibilities as a basis for hope. Were she to cease hoping—were she, that is, to resign herself to her limitations as a dancer—she would not cease to value dancing. Nor would she turn down a Genie's offer to give her the capacity she now lacks. She would, however, no longer regard herself as "falling short." She would be like those of us who would love to have wings, but who understand that to fly is simply not a part of our good. Or like the blind person who never thinks of herself as a sighted person who cannot see. She would no longer be in despair.

The reader who reflects on the histories of human lives—real and imagined—should have little difficulty discovering the paradox of despair on both a large and small scale. I offer a second example of my own in the form of a contrast between two aging women whose youthful beauty has been an important part of their self-conception. Martha looks in the mirror and sees that her face has begun to sag. "I am not as beautiful as I once was," she thinks, with a regretful, but resigned, sigh. Mary looks in the mirror and she, too, sees that her face has begun to sag. "I am not as beautiful as I once was," she thinks with alarm. Neither Martha nor Mary has lost her ability to appreciate human beauty; neither has altered her aesthetic judgments. But whereas Martha believes that youthful beauty is no longer a possibility for her, and so ceases to regard it as to-be-realized by her,[16] Mary believes she can surely look much better than *this*, and makes an appointment with the hairdresser, the make-up artist, the plastic surgeon. She has seen what these wizards can do; so, though she is not happy, she has reason to hope.

---

[15] Note that the impossibility referred to here does not rule out the conceptual possibility discussed in note 24.

[16] Alternatively, we could imagine her as an instance of a common phenomenon to which Sheed calls our attention: "The physical decrepitude of old age seems to bother people surprisingly little. When they were young, the old-timers around here must have dreaded the thought of someday looking the way they do now, all scales and wattles, but when it comes, it must seem like the only way to look"—*In Love with Daylight*, p. 239.

Time goes by. Mary has now tried everything she can think of, and she still falls far short of her ideal of beauty. The evidence is overwhelming: she is not that type of human being any more; she does not "have it in her" to be (that) beautiful; being (that) beautiful is not a possible way for her to flourish; it is not part of her good; it cannot be for her a condition to-be-realized. Still, she is deeply committed to a self-conception which is incompatible with the evidence. In *identifying* herself *as* a beauty, she has *identified with* herself as a beauty. As far as she is concerned, there is no gap between who she is and who she takes herself to be. She thus refuses to give up hope. Unable to deceive herself about the evidence, she nonetheless continues to believe that it is still possible for her to be beautiful in the way only a young person can be. At the same time, she believes (she knows) that this is not possible. So she scrutinizes the mirror in a state of despair. Again, her despair is not simply a function of the fact that she *wants* to be beautiful, nor even of the fact that she wants this *very badly*, where this is a matter of placing it at the top of her preference rankings. She is in despair because she regards (her conception of) human beauty as a part of her good. It is, she believes, a component of her way of flourishing, and so, it is a way she might be—indeed, it is a way she can reasonably expect to be. The paradox is, of course, that she knows better—just as do those who refuse to give up hope despite having been repeatedly rebuffed by their beloved, despite having tried all the infertility treatments many times, despite having lost all enthusiasm for their job a long time ago.

When the conflict between a person's reality and her personal ideals is the result of her own limitations, her contradictory beliefs about her possibilities are essentially contradictory beliefs about who she is. It might seem that the conflict between the real and the ideal does not take this form when it is occasioned by the person's external circumstances. I believe, however, that no such neat contrast can be drawn; and this is because our circumstances are intricately entangled with our conception of who we are. To make this point vivid, I offer one last example. Consider a woman whose only child has just died. Though she believes that her little Johnny is gone forever, she simply cannot believe this. To resign herself to her child's death would be to resign herself to being someone who is not Johnny's mother—someone for whom being Johnny's mother is not a state-to-be-realized by her. So she cannot stop hoping that the whole thing is just a horrible nightmare. So she is condemned to despair.

Let me be clear: the misery in this case is occasioned by the child's death. More importantly, his mother is miserable *about* his death—not about her conflicting sense of who she is. There is, however, a structure to her misery. It is the structure of a conflict between two self-conceptions—a conflict between two conceptions of what is possible, two conceptions of what it is reasonable to expect. This is the structure of despair.

III. THE PERVASIVENESS OF DESPAIR

I believe that unhappiness as despair is a very common form of unhappiness. Indeed, I suspect that all unhappiness, from the most mild to the most intense, has the structure I have attributed to despair. To be unhappy is to be in despair—even if it is not always (or even usually) to be in despair's deepest depths.

This claim may well seem outrageous, vulnerable to powerful challenges from opposite sides. It may seem obvious that neither hope nor despair is a necessary condition of unhappiness. Surely, the alert reader will insist, there are many cases in which someone is in despair, and so is unhappy, precisely because she has given up hope; and surely, people are often unhappy, not because they believe they *cannot* realize some personal ideal, but simply because they believe they *have not* yet realized this ideal (as fully as they would like). I will try to respond to each challenge in turn.

First, then, my conception of despair appears to be vulnerable to any number of counterexamples in which a person is in despair precisely because she lacks all hope. Consider, for example, the case of someone who has just been forced at gun point to leave her burning home and walk for days until she has reached a refugee camp in another country, where she can barely get enough food to sustain herself. Can such a person not reasonably complain that her life falls far short of her ideals, even if she does not believe there to be any real possibility of realizing these ideals? According to this challenge, the possibility to which the refugee is committed in embracing the ideal of returning home and reestablishing a life in which she has plenty to eat is not a "live" possibility, but merely the conceptual possibility of living this way. So her despair at not being able to realize her ideal is not paradoxical, after all: she has no hope of ever "starting over" again.

In this account, the role I have assigned to the attitude of hope is replaced with something like a mere wish. I want to argue, however, that this substitution does not work: the negative condition constituted by an unfulfilled wish is not the negative experience of being unhappy; if someone merely *wishes* that a given state of affairs would obtain, then her negative attitude toward the status quo does not include the disappointed or frustrated *expectations* that are essential ingredients of unhappiness.

Let us take a closer look at the refugee's wish-like attitude toward her ideal. By stipulation, it reflects the belief that it is unreasonable to expect the ideal to be realized. Yet in order to inspire despair, this attitude cannot be an "idle" wish of the sort one might direct toward a state of affairs that one likes very, very much but believes cannot obtain outside the realm of fantasy. That is, it cannot be an idle wish, even though it must be a wish she believes can never be granted. But what sort of attitude could possibly satisfy this requirement? Only, I submit, a paradoxical one—only an attitude whose object is something one regards as to-be-achieved, even though one believes that it really can*not* be achieved. If one's belief that one cannot achieve *X* is a source of despair, this must be because *X* is one's end. And if *X* is one's end, then one must believe that it is possible to achieve it.

Despair would not be paradoxical if it were possible to be unhappy about one's failure to satisfy some desirable condition without regarding this condition as (more fully) realizable, and so, as an end. It might seem that nothing could be easier. After all, can one not believe that things ought to be a certain way, while recognizing that they cannot possibly be this way? The answer is that one can, indeed, do this, but that unless, in addition to believing that one's life ought to be a certain way, one also insists, or demands, that it be as it ought, one has insufficient psychological material for despair. One can be quite "philosophical"—even complacent, bored, or amused—at the recognition of yet another illustration of the fact that "ought to be" does not imply "can be." To be sure, one may be surprised that the two are out of sync in this particular case; one may wish that they were not. But such attitudes do not suffice for despair, as is evident from the fact that most people manage to "grow up" without being laid low by the inevitable discoveries of the many ways in which real life falls short of what it really ought to be.

Think, for a moment, of how many aspects of our world are not as you believe they should be. Surely, you do not like this fact. You would probably be willing to call it an "unhappy fact," and to express your view by saying "I am *not happy with* the way things are (in, for example, the slums of Calcutta)." But this attitude is not unhappiness. Your recognition of the gap between what is and what ought to be does not perturb the waters of your affective life; it is not the source of your "ups and downs." Or rather, when such gaps do "get you down," this is not simply because things are not as they ought to be, but because they are not as they *must* be, given your goals.[17]

Unhappiness comes on the scene only when one actually insists that reality (one's real life) be as it ought to be.[18] One may address one's demand exclusively to one's pillow at night, or—silently—to nothing, and no one, in particular. What matters is the demanding attitude itself. Unless, moreover, one believes that it is reasonable to expect one's demands to be satisfied, one cannot believe that it is reasonable to make them. So, since one necessarily takes it for granted that one's demands are reasonable (at least if one is both sincere and sane), it is part of the very point of one's demands that one is reasonable to expect them to be met. We thus arrive, again, at the paradox of despair: despair at the unrealizability of one's ideals presupposes the belief in their realizability; in despairing that one's demands cannot be met, even while believing that it is reasonable to make them, one reveals the hope at the heart of one's ideals, and so one reveals that one has not yet really given up hope after all.

It is important to stress, again, that the relevant demand that *X*, *Y*, or *Z* happen is not merely the claim that this is what "*justice* demands." Justice demands that something be done about the slums in Calcutta. But one can believe this—believe it fervently—without feeling unhappy. In order to be unhappy about something one does not like, one must not only prefer that it be otherwise, but this change must be something one aims at; it must not only appear to be desirable; one must embrace it as a goal to-be-achieved—not just someone else's goal, but one's own. Unhappiness is essentially paradoxical precisely because it is essentially the experience of having a goal that one cannot see one's way to (coming closer to) achieving. Having the goal requires hope; seeing no way to achieve this goal—or even to come closer to achieving it—inspires despair. In most animals the force of instinct ensures that they have no goals they regard as impossible to achieve. We are made differently, however; and our capacity for unhappiness is one of the consequences.

---

[17] These are, of course, not the only possible attitudes along a continuum. In "feeling sorry" for another person, for example, one may experience an intense pain which is as distinct from feeling unhappy as is the pain of a sprained ankle. When the person for whom things are not going well is one's own child, one's sympathetic response can easily metamorphose into unhappiness. This is because one "identifies with" her. Her hopes are one's own, and—however foolish this may be—one has hopes (not just dreams) *for* her. That is, among *one's own* possibilities are ways *she* can be: one is definitely *not* the mother of someone who flunks out of high school, or robs a bank!

[18] Again, this is Kant's point about regulative ideals. As Susan Neiman explains, "The positing of an end is equivalent to a demand for its realization, and ideas of reason simply are ends"—*The Unity of Reason: Rereading Kant* (New York: Oxford, 1994), p. 69. "Regulative principles...[are] simultaneously ideas of and motives for the realization of, a certain possiblity" (p. 89). Kant calls them sources of "guidance"—*Critique of Pure Reason,* A827/B855, p. 649.

Many will grant that despair is closely related to hope. But according to conventional wisdom, the two attitudes cannot be simultaneous if they are directed toward the same thing: despair is either the loss of hope or the attitude that supplants hope when it is lost.[19] If I am right, the conventional wisdom is mistaken, for it fails to do justice to the difference between ceasing to believe that something desirable is possible and experiencing this shift as a loss. To experience the impossibility as a loss, one must still regard the unsatisfiable condition as a condition that is not only worth satisfying, but to-be-satisfied, an end, a goal, a live possibility even still. There is another bit of conventional wisdom that almost acknowledges this paradoxical attitude. “She simply cannot reconcile herself to her situation,” we say. In other words, though she has the situation vividly, and painfully, before her mind’s eye, some part of her refuses to regard it as an accurate picture of reality. (Consider the way that despair tickles the edges of panic in the thought of the frantic (*desperate*) parent who rushes about in search of a child lost at the zoo: “What cannot be might nonetheless be.” It *cannot* be because it *must not* be. And the demand does not change when “might be” changes horribly to “is.”)

Insofar, then, as our personal ideals are not merely the object of fantasy—insofar as they are not merely the focus of our most heartfelt wishes—they resemble our regulative ideals in presupposing hope. They presuppose the hope that they will be realized; for without this hope, they could not function as ideals—goods to be realized by us. This means that there is something intrinsically paradoxical about being unhappy about the unrealizability of one’s ideals.

I believe that all unhappiness takes this form. For, again, I cannot see how there could be negative emotional states that deserve to be called “unhappiness” despite lacking the goal-oriented structure I have tried to describe; I cannot make sense of the idea that someone could be unhappy about something (rather than merely wishing it were otherwise) if her unhappiness had nothing to do with her expectations. Perhaps I am wrong about this. Perhaps the sorrow we feel when a loved one has died, for example, has an essentially passive component, as well as the active component I have been stressing here.[20] To defend this possibility, however, one cannot merely insist that true love *just is* an emotion or attitude which can be manifested by a grief that has nothing to do with dashed expectations and unrealized ideals. One must make sense of this unhappiness; one must explain how it is, essentially, a manifestation of love.

---

[19] In analyzing hope, J.P. Day succinctly expresses the conventional wisdom on the subject: “A cannot despair that P at the same time as he hopes that P...”—“Hope,” *American Philosophical Quarterly*, VI (April 1969): 89–102, here p. 94.

Even if this explanatory burden can be met—even if there are aspects of a person's perceived good that are independent of her ideals, and so lack the practical structure of these ideals—our conception of our good is generally inseparable from our ideals. In most cases of unhappiness our unrealized good thus functions for us as a practical guide, something in relation to which we orient ourselves. Accordingly, paradox is not simply a feature of the occasional weak-willed act, and of the more frequent episodes of self-deception. Paradox is at the center of one of the most ordinary experiences of our emotional life.

But what about that second challenge to my account—the challenge from the other side? Even if unhappiness presupposes hope, why should we think that it presupposes despair? Surely, there are many cases in which a person is unhappy about gaps between the ideal and the real that, she thinks, are difficult, but not impossible, to close. Could anything be more obvious?

In fact, however, the appearances are misleading. Wherever we seem to see unhappiness without despair, we are really seeing either someone who wishes that things were different but would rather not do what it takes to change them, or someone who is in despair, after all. Let me explain.

Someone might wish that the gap between her personal ideals and her reality were not so big. Suppose, however, that she does nothing to close this gap. This could be for one of two reasons: either she believes that she *cannot* close it, or she does not *want* to close it, all things considered. Since in the former case, she despairs of closing it, it is the latter case that interests us here. If she does not want to close the gap, this must be because she believes that the cost of doing so is too great—that realizing (or more nearly realizing) the relevant ideals would involve foregoing other good things which, all things considered, it is better not to forego.[21] But if this is her situation, then she is not in fact unhappy, all things considered, about the fact that she has failed to make the effort necessary to close—or at least to narrow—the gap between her reality and her ideals.

---

[20] I am grateful to Maggie Little for pressing me on this point.

[21] Of course, if it would, in fact, be relatively easy for her to bring reality into greater harmony with her ideals, then there is good reason to suspect that they are not really her ideals, after all. We are to imagine a case in which the costs are high enough to make it quite difficult to pay them, but not so high as to induce despair.

Of course, she may well feel unhappy about the fact that she finds herself in a situation in which it would be better not to make the effort. But if this is the case, then the object of her unhappiness must be the high price of closing the gap; for otherwise she would not be unhappy about the fact that the unrealized ideals are unrealized. If she is unhappy about the fact that her ideals are unrealized, then she must be unhappy about the fact that she would have had to pay a very high price to realize them. But this is the unhappiness of despair: she is unhappy that she cannot both realize her ideals and preserve the other good things whose sacrifice constitutes the high price she would have to pay in order to do so. In other words, she is unhappy about the fact that she cannot have two things at once, each of which is, she thinks, essential to her living a good life. She is unhappy because she does not believe that it is possible to realize this ideal condition.

IV. THE IRRATIONALITY OF UNHAPPINESS

Having discovered the paradoxical structure of unhappiness, we are in a position to recognize the principle of rationality that opposes it. When we are unhappy about the unrealizability of our ideals, we hold self-contradictory beliefs. And when we do this, we violate a fundamental principle of theoretical reason. Since our reason is hostile to contradictions, it tells us to stop despairing. It tells us to stop being unhappy—and to avoid future unhappiness, too.

The point is not that we are especially rational when we manifest our natural tendency to react positively to a perceived harmony between our reality and our ideals. To say that our reason supports our self-concern is *not* to say that it tells us to be preoccupied with ourselves, and with our own happiness, in particular. For all our reason cares, we can be so caught up in our "projects," so immersed in the business of realizing our ideals and maintaining the bare necessities of life, that we do not have a spare moment to consider the relationship between the way things are and the way we believe it would be ideal for them to be, nor even half a spare moment to ask ourselves how we feel about this relationship.[22] Only when we do, in fact, feel unhappy does our reason reinforce our natural desire to escape this unpleasant condition. And again, this does not mean that, *all things considered*, we ought to do what we can to escape it, but only that one of the things to consider is the fact that we cannot remain in this condition without contravening a formal principle of rationality. In short, we have *a* reason to try to be happy. Without the slightest allusion to other rational agents, or to the objective value of happiness, our reason supports our self-concern.

---

[22] I would like to thank Martha Nussbaum for forcing me to clarify this point.

There are several ways to obey reason's demand to cease holding contradictory beliefs about our personal ideals. One way is to give up the ideals, as human beings often do when they make the transition from the innocence of childhood to the experience of adulthood. Our ideals would not be worth much, however, if we could give them up whenever they proved "unrealistic." And as I noted earlier, *we* would not be worth much either if we lacked the loyalty, constancy, and courage that holding on to our ideals sometimes requires. This is why the hope at the heart of despair is not wholly irrational. Or perhaps we should say that there are good reasons for some irrational hopes, reasons that pure reason does not understand. (There is sometimes good reason to cry over spilled milk. Better: when what is spilled is very precious to us, there is often good reason to cry—and cry, and cry.)[23]

If someone were not made unhappy by the death of her child, we would think there was something terribly wrong with her. We would not normally express this thought by saying that she has sufficient reason to be unhappy. But we could: she has reason to be unhappy because he is dead; his being dead is a reason—a sufficient reason—for her to be as unhappy as a human being can possibly be. Implicit in this verdict is an ideal of humanity: human beings with "any feeling" would be devastated by this event; and "having feelings" is something we take to be part of any minimally decent human life.

For most of us, this ideal of humanity is a personal ideal. And most of us are incapable of failing to realize it: no matter how hard we might try, we could not purge ourselves of the wide range of emotions that characterize our daily experience. (See, however, the following discussion of resignation.) Given our ideals, we rightly regard the naturalness of our feelings as a fortunate fact. Nonetheless, our natural susceptibility to unhappiness makes us vulnerable to irrationality. Indeed, if I am right, it *is*, essentially, a natural susceptibility to irrationality.

---

[23] Derek Parfit declares: "Grief isn't irrational simply because it brings unhappiness. To the claim 'Your sorrow is fruitless,' Hume replied, 'very true, and for that very reason I am sorry' "—*Reasons and Persons* (New York: Oxford, 1984), p. 169. "We may," Parfit writes, "object to a world in which our loved ones are taken away, but if they are taken away, we do not want to fail to experience the fact, to register it as an evil." I hope it is clear that my agreement with Parfit is far greater than my disagreement. On my account, unhappiness is not irrational because it is fruitless. Rather, it is irrational because it includes the belief in the fruitlessness of hoping that things will be different. What is more, the irrationality at the heart of unhappiness is only *a* reason to avoid it. As I have stressed, there are often other, even better, reasons to be unhappy.

Human life is such that for every one of us (if we live long enough), there will be occasions on which we cannot realize the aforementioned ideal of humanity without having self-contradictory beliefs—without, that is, experiencing one of our other ideals as a good it is both reasonable and unreasonable to expect ourselves to realize. As I suggested earlier, on these and other occasions, self-contradiction is also the price we must pay for realizing the meta-ideal of being "true to our ideals." So, though on all such occasions we have good reason for reacting as we do, we always also have a reason *not* to react this way: we would thereby avoid self-contradiction. Since the ideal of rationality is another of our personal ideals, the pressure to "overcome" our humanity, and to "give up" ("give up on") our old conception of what counts as a good life for us, is pressure we impose on ourselves.

Fortunately, we need not always give up our ideals in order to satisfy reason's demands to avoid the self-contradiction of unhappiness. Sometimes we have another option: we can do something to prove to ourselves that the evidence against the possibility of realizing an ideal is not as conclusive as it seems—that we are not really unreasonable to expect the ideal to be realized. Under the pressure of theoretical reason, someone may finally screw up the courage to quit her high-powered job, and seek less prestigious (and less lucrative) employment that will leave her some time to devote to her poetry. The evidence against her ability to "make it" as a poet is, she thinks, overwhelming. And she has been dragging herself through her days with a barely acknowledged despair at ever being the poet she still believes she might be. It has taken a while. But since she cannot bear to abandon hope, she has finally come to see that her only rational option is to take the steps necessary to establish that the evidence against her has been misleading.

This woman is lucky: it is possible for her to do something to try to justify her ideal. Many others are not so fortunate, however. Short of deceiving herself, for example, there is nothing the aggrieved mother can do to convince herself that her child is not really dead. In her case, theoretical reason thus gives her no choice: it commands her to give up her hope. Again, it may be that she cannot, or will not, obey. She may rightly sense that despair is her most honorable option, that she would be a traitor to what matters most if she were any less unhappy and broken.[24] It is an interesting question under what conditions a despair that was initially justified becomes unreasonable; but I will not try to answer this question here. In most cases, it seems, the desperate hope gradually atrophies, to be replaced by the resignation that lies in the vast no-man's-land between happiness and unhappiness. There are other cases, however, in which the painful modification of one's ideals yields a joyful affirmation of one's new self-conception, and of the new life that goes with it. Think, for example, of someone who has despaired for years over his homosexuality, and who finally adjusts his ideals to fit the contours of his possibilities as he perceives them.[25] Sartre is right to note how often we cling to our self-conception as a convenient excuse. But he is wrong to suggest that someone is necessarily in bad faith if his sense of his possibilities reflects his belief that he is a type of human being whose good differs in important ways from the good of most others.

---

[24] An ideal generally combines several valued things into one. Accordingly, it is rare that we must abandon an ideal entirely in order to avoid irrationality. As Little has reminded me, we can usually revise our ideals in such a way that aspects of the old remain in the new. Thus, the would-be dancer might strive to become a choreographer; the wheelchair-bound man who had been training to race the mile might strive to become the world's fastest paraplegic.

Recently, some philosophers have suggested that we cannot betray our deepest values and ideals without destroying ourselves. Thus, according to Frankfurt, "Agamemnon at Aulis is destroyed by an inescapable conflict between two equally defining elements in his own nature: his love for his daughter and his love for the army he commands.... When he is forced to sacrifice one of these, he is thereby forced to betray himself. Rarely, if ever, do tragedies of this sort have sequels. Since the volitional unity of the tragic hero has been irreparably ruptured, there is a sense in which the person he had been no longer exists"—"Autonomy, Necessity, and Love," in his *Necessity, Volition, and Love*, pp. 129–41, here p. 139, n. 8. Similarly, Christine Korsgaard argues that "to violate [the conceptions of yourself that are most important to you] is to lose your integrity and so your identity, and to no longer be who you are. That is, it is to no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living and your actions to be worth undertaking. It is to be for all practical purposes dead or worse than dead"—*The Sources of Normativity* (New York: Cambridge, 1996), p. 102.

I think David Velleman is right to criticize such claims as obsuring the distinction between a person's self-conception and her identity (her "self")—between self-betrayal and suicide—see his "Identification and Identity," in Sarah Buss and Lee Overton, eds., *Contours of Agency* (Cambridge: MIT, 2002), pp. 97–100. A severe blow to one's self-conception may be worse than death, but it is not the same thing: one can suffer such blows without ceasing to exist. (Think, for example, of what happens to Lord Jim at the beginning of Conrad's novel—and of how he, the same Lord Jim, responds to having done what he knows that he—Lord Jim—would *never* do, and so could not possibly have done.)

An appreciation of the fact that we can survive profound changes in our self-conception should warn us against mistaking the *contingent*, though very real, limitations on our possibilities for *conceptual* limitations that we could not, even in principle, exceed without ceasing to be. In other words, though a person's good clearly depends on who and what she is, it is conceptually possible for a person to change so much that her good changes too. Indeed, for all we have reason to believe about the conditions of personal identity, a person could persist as a member of another species; and so, for all we know, she could, in principle, acquire the features that would make flying an essential component of her good. Aristotle is right when he reminds us that "no one [who is rational] chooses to possess the whole world if he has first to become someone else...; he wishes for this only on condition of being whatever he is"—*Nicomachean Ethics*, W.D. Ross, trans., in *The Basic Works of Aristotle*, Richard McKeon, ed. (New York: Random House, 1941), 1166A19, p. 1081. But it does not follow, as Nussbaum suggests in commenting on this passage, that what could possibly be a good for us is limited by our imagination. Even if a life "was so remote from mine that I could not imagine in it a person whom I could accept as identical with myself"—"Aristotle on Human Nature and the Foundation of Ethics," in J.E.J. Altham and Ross Harrison, eds., *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (New York: Cambridge, 1995), pp. 86–131, here p. 91—it might nonetheless be a life I could live. If this possibility exceeds the limits of my imagination, then it is not relevant to my ideals, and so, as Korsgaard and Frankfurt remind us, it cannot constrain my choices; it is nothing I must, or can, be "true to"

V. RESIGNATION

The gay man who eventually embraces his homosexuality and the bereft mother who eventually "resigns" herself to her child's death illustrate two very different ways of adjusting to conditions that limit the sort of life one can live: whereas he no longer desires to be heterosexual, she would give anything to have her son back. Yet, I said, this is not a contrast between a happy and an unhappy person. Though the woman is anything but overjoyed about her new self-conception, she is resigned to it. That is, she has no hope that she will be someone's mother again, and so no despairing, disappointed hope—no despair, no disappointment, no unhappiness.

Resignation is the absence of hope without paradox. In some forms it is benign; in others it approaches pathology. In all it is a no-man's-land between happiness and unhappiness. Consider, again, the woman whose child has died. She may be resigned to the fact that she cannot drink water while singing an aria; she may be resigned to the fact that she experiences monthly pain and discomfort when she menstruates; and she may (eventually) be resigned to the fact that

---

in deciding what to do. We can concede all this, however, while at the same time acknowledging that I could, conceivably, acquire ideals in the future which would require such a great self-transformation that I cannot now imagine surviving it.

[25] The account offered here sheds light on a basic feature of happiness to which Charles Larmore calls our attention in a recent article: "We are never in a position to grasp in advance the full character of our good, even in its broad outline. As a result, our happiness includes not just the anticipated good we achieve, but also the unexpected good which happens to us"—"The Idea of a Life Plan," *Social Philosophy and Policy*, XVI, 1 (Winter 1999): 96–112, here p. 103.

her child is dead. If her attitude is the same in all of these cases, it nonetheless takes different forms. To say that she is "resigned" to the first condition is to suggest that she accepts it—and the limits it imposes on her possibilities—without giving it any thought. This could also be her attitude toward the second condition; or—alternatively—she could frequently note how nice it would be if things were otherwise.[26] Finally, if she really does manage to resign herself to her child's death (and many in her situation never do), then this would almost certainly at least initially involve an emotional/psychological numbness in response to everything having anything to do with him, and with the awful event.

Without straining ordinary language we can, if we want, characterize this last form of resignation as a different sort of despair: despair without anguish—indeed, without affect of any kind. When it is directed toward only one dimension of one's life, it creates a barricade against the entrance of hope along this dimension. When it is directed toward one's life as a whole, it is a kind of death.

Suicide is an extreme response to the belief that there is "nothing to live for," one's hopes are hopeless. But one can kill one's *hopes* without ending the life of one's body. One can abandon hope altogether by ceasing to have any ideals. Someone in this condition is psychically/emotionally numb. She has no conception of her possibilities. So she has no conception of who she is. So she is dead to herself. Like all who are dead, she is invulnerable to disappointment. She is not happy, of course. But she is not unhappy either.[27]

---

[26] Notice how different this attitude is from the unhappiness of disappointed expectations. As this example reminds us, there would be far more misery in the world if Freud had been right—if, that is, little girls really did conceive of themselves as deformed boys. At some point in their lives half the human population would suffer in much the way that many homosexuals suffer in our society today. Perhaps there will come a time when we will be able to laugh at the thought that human beings must overcome painful disappointment in order to accept their erotic longings for others of the same sex, just as we can now laugh at the thought that some (most!) human beings must overcome painful disappointment in order to accept the fact that they do not have a penis.

[27] Someone who is profoundly depressed might seem to fit this description: she is too depressed to hope for anything because she is too depressed to care about anything. In most cases of serious clinical depression, however, the person is in such great pain that she desperately hopes the pain will go away. This motive for suicide is distinct from the "anti-motive" of having nothing to live for, though clearly, the latter is a source of the former; and the power of the former reinforces the latter. Following the next paragraph in the text, I call attention to how extremely difficult it is for us to reduce the intensity of certain basic forms of suffering by simply picking new targets for our hope.

Because hopeless despair is affectless despair, it is easy not to notice that one is in this condition. For a powerful description of this numb state of self-alienation, consider the following poem by Emily Dickinson: "There is a pain—so utter—/ It

Far better to be irrational than to lack the capacity for happiness. Far better to be unhappy. But when some measure of happiness can be purchased by killing off some part of oneself, the price may be worth it. There may be no other way to "get on with one's life." Of course, in most such cases, what is dead is not really gone. So there is usually room in such lives for a special hope. Someone whose child has died may eventually dare to conceive of herself as someone who can think again with pleasure of the many wonderful moments she spent with him. If, at the same time, she is convinced that this can never be, then her hope will be a new source of despair. If she holds on long enough, however, she may one day realize her ideal: she may succeed in sustaining the absence of any hope that things will be as they once were, without sustaining an absence of feeling.[28]

And what of the hungry, frightened refugee? Can she ever truly resign herself to her lot? It does seem that others in her situation have done just this. But I wonder whether such apparent resignation is not really just a case of biding one's time—waiting for things to get better. Waiting resignedly, perhaps. But waiting, just the same. Hoping.

The difficulty of giving up hope in such a case is not merely psychological; it is a function of the basic animal drive to thrive, and of the fact that there are limits to the conditions under which human animals can thrive—limits imposed by their need to eat and drink, by their sensitivity to pain, and so forth. Like all animals, we humans have the "built-in" goal to thrive in the ways that are characteristic of our species. And this goal is naturally part of our conception of who we are: I am someone who does not live on roots and leaves, who drinks water (or some other liquid), who sleeps daily. As my earlier allusion to blindness suggests, our self-conceptions can leave out many of the features that characterize a normal healthy human animal: one can conceive of oneself as the kind of human being who does not see, or whose legs do not move, or who does not make insulin. If, however, my happiness required me to embrace a conception of myself as the kind of human being who is perpetually cold and hungry, then it

---

swallows substance up—/ Then covers the Abyss with Trance—/ So Memory can step/ Around—across—upon it—/ As one within a Swoon—/ Goes safely—where an open eye/ Would drop Him—Bone by Bone," from *The Poems of Emily Dickinson*, Thomas H. Johnson, ed. (Cambridge: Harvard, 1979), Volume II, p. 460.

[28] Emotions come and go. It is thus highly likely that someone who has suffered such a profound loss will become unhappy again for certain periods of time. Upon looking over a photo album, for example, she might relive her loss as though it were just yesterday; whatever she thought and felt when she first heard the news would "all come back again."

would probably be impossible for me to be happy. My animal nature would continue to remind me that I really do regard a full belly as part of my good. It would not permit me to give up the aim of eating a decent meal. It would keep me hoping; and so it would sentence me to despair.

Resignation is not alchemy: it cannot turn bad states of affairs into good ones. Nonetheless, it is often the necessary first step in revising one's conception of one's good.[29] And so, it is often the first step—a rationally justified step—on the way back to happiness. Again, most such revisions do not depend on concluding that the grapes were probably sour anyway; for, again, someone can readily admit that she would welcome a change which she no longer hopes for—and even a change which she no longer takes to be very important. In some cases, however, when, like the transformed homosexual, a person concludes that he was wrong about his own good, the alteration in his personal ideals is inseparable from an alteration in his more general views about what is desirable. Suppose, to take another example, that Bob has believed since childhood that no human life is a good human life unless the human being who lives it has been "saved." And suppose that he despairs of ever achieving salvation himself. One way for him to alleviate his unhappiness would be for him to abandon his belief. He might not be able to do this, of course. But what if he could? What might this abandonment look like? He might conclude that there is no God, after all. Or he might conclude that he, at any rate, has no God. In either case, if he succeeds in becoming a nonbeliever, he will no longer value salvation. Indeed, he probably would not even recognize it if it hit him on the head. (This is the difference between a true nonbeliever and someone who has turned his back on God in despair.[30])

---

[29] Spinoza, Rousseau, and many others have tried to impress upon us how much our happiness depends on our ability to recognize and accept what we cannot change. This article can be read, in large part, as an exploration of this Stoic insight, and an attempt to show that the connection between refusing to accept what is necessary and being unhappy is a more intimate connection than that between cause and effect. It is interesting to compare my remarks about resignation with the following comment by Allan Bloom: "Irony flourishes on the disproportion between the way things are and the way they should be while accepting the necessity of this disproportion. It is a classical style because the ancients did not expect that reality could become rational...moderation, rather than being the expression of a timid or easygoing soul, was for them the expression of one who has overcome hope and therefore indignation.... In short, irony seems to presuppose the distinction between theory and practice..."—*Love and Friendship* (New York: Simon and Schuster, 1993), p. 193.

[30] My claim about despair thus represents a challenge to the teaching that despair is the worst sin because it involves giving up on God.

One can sometimes deceive oneself into believing that one's conception of one's good life has really changed—and with it, one's conception of oneself as someone for whom it is good to live such a life. Theoretical reason gives mixed messages about such self-deception. On the one hand, it tells us not to permit our wishes to determine our conception of reality. On the other hand, it tells us to avoid self-contradiction. I suspect that if deceiving oneself about what one really deems valuable and important does not actually cause one to change one's mind, then this sort of self-deception is especially difficult to maintain. No sooner is the grape in one's mouth than one concedes that life was no good without it; one is overwhelmed with relief and joy when God finally shows His face.

It is often difficult, too, to deceive oneself about one's real possibilities. But I suspect that many happy people whose ideals are far from being realized owe their happiness, at least in part, to their ability to induce in themselves the false belief that the evidence does really warrant their hope. In some of these cases, the self-deceived belief becomes self-fulfilling. In all cases, it is an antidote to despair.[31]

When we comply with reason's demand by deceiving ourselves in this second way—and especially when we are incapable of pursuing the other ways of complying—we are, it seems to me, in a condition that is as close to Kantian "rational faith" as it is possible to come in relation to one's personal ideals: we have embraced a "useful fiction" which has the support of a fundamental principle of reason, even though the evidence tells strongly against it. In this case, of course, reason's recommendation is conditional, in part, on contingent psychological factors; reason is not, itself, the source of the ideals. But nor is it, on the Kantian account, the source of the human desire for happiness, without which happiness would not be part of the Highest

---

[31] Sheed offers an autobiographical example of how deceiving oneself can contribute to the gradual process of reconceptualizing oneself in the way one must in order to be happy: "Nobody ever got used to blindness, or living in a wheelchair, or crawling to the bathroom, in his mind; yet when the time comes, the body often moves amazingly fast, handing out instructions as it goes on how to handle the latest disaster, like the next clue in a game. Your task is to keep your mind occupied between clues, keeping the truth or at least the words at bay until your body has learned the next moves. If, for instance, I'd been told at the beginning exactly how far I'd get with my exercises, and where I'd end up, I might indeed have felt like a cripple and quite bowed down by it; but I was able to convince myself that I was going all the way until my body was ready to tell me otherwise, and until I'd learned slowly for myself, pushing against that unyielding hand, that this is where you get off.... [A] firm belief in miracles and the power of prayer, and a trusting nature, did treat me to three and a half years of unbridled optimism, at the end of which I was so used to polio that I could barely remember what life had felt like without it"—*In Love with Daylight*, pp. 31–32.

Good. If, as I have suggested, our happiness is tied to our perception of the relationship between our actual condition and our personal ideals, then the useful fiction that the gap between our real life and our ideal life is not so big—or if big, then at least not so difficult to close—may play an important role in sustaining our rationally mandated faith in the possibility that someday everyone who ought to be happy will be.

VI. THE RATIONALITY OF OPPRESSION

Many arguments with other rational agents about what each has reason to expect of the others are, in a broad sense, *political* arguments. Before bringing this article to a close, I want to say a few words about the political implications of the irrationality of unhappiness. Let me begin by stressing what my thesis does *not* imply. It does *not* imply that you can deprive me of something essential to my good—that, for example, you can deny me the right to vote—and then declare that I am irrational to protest, since what you have taken away is no longer attainable. I can reasonably retort that it is *not* irrational for me to protest, since it (the right) *is* attainable; it is attainable, since you can easily change things so as to enable me to vote. In short, you can deny me the right to vote without forcing me to conclude that it is unreasonable for me to expect that you will grant me this right.

This having been said, your behavior (and that of your associates in power) may well convince me that I have no reason to expect that you will grant me the right any time *soon.* In this case, if I continue to demand the right *now,* and if this is not simply a matter of strategy (or rhetoric), I will almost surely be unhappy. ("*Almost* surely": unhappiness *requires* a tension in one's beliefs about what is possible; but again, though it is our natural way of experiencing this tension, I do not see any reason to insist that it is the inevitable consequence.) If I am unhappy (as opposed to merely judging that things are not as I would like them to be, nor as they should be), I will be (in the sense that we have been exploring) irrational: since I do not believe that it is possible to gain the right soon, I would be more rational if, rather than clutching at a desperate hope, I did not "let it bother me" that I will have to wait.

Again, this merely implies that I have *a reason* to cheer up—not that, *all things considered,* I *ought* to do so. Happiness is not the only thing of importance in life. Certainly, rationality is not. Furthermore, in changing my attitude, I need not alter my long-term expectations. And so I need not give up my belief that I will, eventually, be granted the right. Nor need I give up fighting for this right. Indeed, my hopeful long-term expectations will surely depend on my belief that I will never give up.

What if, however, your behavior (and that of those with whom you share power) convinces me that this belief, too, is unfounded—that there is no reason to hope I will ever be granted the right to vote (or to drive a car, or to attend university, or...)? Unfortunately, we are all too familiar with situations of this kind, and we know the hard choice that they offer: like so many other victims of oppression, I will have to choose between being unhappy for the rest of my life and being happy with less than I had hoped for. If I really have good reason to believe that these are my only alternatives where my happiness is concerned, and if it really is within my power to choose, then, it seems to me, I would not necessarily be wrong to opt for happiness. And, in any case, I would not be irrational to do so.

Of course, this is just what you, my oppressor, are counting on me to recognize. You are counting on me to try not to let my condition bother me (not to let it "get me down"), even if I dislike it; and you know that if it stops bothering me (if it stops "getting me down"), then I am eventually likely to take your restrictions for granted when I am thinking about who I am and what is essential to my own good—or if I will not ever reach this point myself, then my children will (or those of my sex, if the restrictions apply only to them). This outcome is compatible with my continuing to fight on behalf of the generations to come. But when I am no longer unhappy, the motivation to fight will be hard to sustain. If I see this subtler side to oppression clearly enough, I will recognize an additional reason to opt for unhappiness: I do not want to be a collaborator; I *am not* a collaborator.

Or am I? After all, I am certainly someone who wants to be happy. I hope that I will one day be happy. Can I really reconcile myself to regarding happiness as a very nice condition which is, however, not-to-be-achieved by me? And if I can, should I? What it is rational for me to demand of you (and of all other rational agents) depends on what it is rational for me to demand of myself. I may be confused about this. And since my confusion will make it more difficult for me to resist you, your success in oppressing me will in this case, too, be due, in part, to the irrationality of unhappiness.

There is considerable philosophical interest in the fact that we sometimes have *practical* reasons for holding beliefs that are not well supported by the evidence. In seeking a justification for our desire to be happy, I have discovered a simple *theoretical* reason for satisfying this desire.[32] It turns out that the impulse to bring our lives into line with our ideals and the desire to bolster our confidence in the appropriateness of these ideals are not really as distinct as we have been inclined to think.

With this discovery, we have, in effect, moved from Kant to Hegel. At least I think this is part of what Hegel was getting at when he told his magnificent story about how an ill-founded confidence in our conception of ourselves and our world generates a desire to *transform* ourselves and our world. Hegel to the contrary notwithstanding, the Real is not the Ideal. Nonetheless, we are committed to bringing them closer together. And we honor this commitment when we pursue our own happiness.

This is not to deny, of course, that we are also committed to acting from motives that we can justify to others. Our reason has heterogeneous ends; and nothing I have said by way of justifying our self-concern implies that this justification is ever overriding. It may well be that we should act on the desire to be happy only in those cases in which our behavior could be endorsed by all other rational agents. Nonetheless, if I am right, rational agents necessarily endorse the desire itself; this is part of what it is to be rational.

According to T.M. Scanlon, "individual well-being [is] morally significant...not because it is intrinsically valuable or because promoting it is self-evidently a right-making characteristic, but simply because an individual could reasonably reject a form of argument that gave his well-being no weight."[33] I have tried to show that one reason why arguments can reasonably be rejected if they give an individual's *happiness* no weight is precisely because, at least in most cases, unhappiness is intrinsically irrational; it is intrinsically irrational because it is itself a form of irrationality.

SARAH BUSS

The University of Iowa

---

[32] According to Darwall, "practical reason includes no intrinsic requirement that we care either about others or about ourselves, since rational agency seems possible without even the capacity to care about a person, oneself or another, for that person's sake"—"Self-interest and Self-concern," pp. 169–70. I am inclined to disagree. My point, however, is that even if Darwall is correct, reason in its *theoretical* capacity does support (at least one form of) our self-concern.

[33] "Contractualism and Utilitarianism," in Amartya Sen and Bernard Williams, eds., *Utilitarianism and Beyond* (New York: Cambridge, 1982), pp. 103–28, here p. 119.

COMMENTS AND CRITICISM

ON FLOYD AND PUTNAM ON WITTGENSTEIN ON GÖDEL

In a recent discussion piece, Juliet Floyd and Hilary Putnam[1] present a new analysis of Wittgenstein's "notorious paragraph" on Gödel's first incompleteness theorem. Textually, they claim that Wittgenstein's remarks have been widely misunderstood, and they argue that Wittgenstein had a better understanding of Gödel's theorem than he has often been credited with. Substantively, they find in Wittgenstein's remarks "a philosophical claim of great interest," and they argue that, when this claim is properly assessed, it helps to vindicate some of Wittgenstein's broader views on Gödel's theorem.

Here, I address the second of these two arguments (while disclaiming the scholarly credentials required to assess the first, the purely textual, one).[2] I begin by examining the central claim which Floyd and Putnam attribute to Wittgenstein, showing that their argument for this claim is inadequate and that the claim itself is almost certainly false. I then argue that, even if Wittgenstein's central claim *were* true, it would not lead to the conclusions Floyd and Putnam think it does. At the end of the day, I conclude that Floyd and Putnam provide no new insights into Gödel's theorem by way of their reading of Wittgenstein.

I. ON FLOYD AND PUTNAM

Let us begin with the relevant passage from Wittgenstein.

> I imagine someone asking my advice; he says: "I have constructed a proposition (I will use 'P' to designate it) in Russell's symbolism, and by means of certain definitions and transformations it can be so interpreted that it says: 'P is not provable in Russell's system'. Must I not say that this proposition on the one hand is true, and on the other hand unprovable? For suppose it were false; then it is true that it is provable.

---

[1] "A Note on Wittgenstein's 'Notorious Paragraph' about the Gödel Theorem," this JOURNAL, XCVII, 11 (November 2000): 624–32.

[2] For more on the textual questions, see Floyd, "On Saying What You Really Want to Say: Wittgenstein, Gödel, and the Trisection of the Angle," in Jaakko Hintikka, ed., *From Dedekind to Gödel: Essays on the Development of the Foundations of Mathematics* (Boston: Kluwer, 1995), pp. 373–425; Floyd, "Prose versus Proof: Wittgenstein on Gödel, Tarski, and Truth," *Philosophia Mathematica*, IX (2001): 280–307; and Mark Steiner, "Wittgenstein as His Own Worst Enemy: The Case of Gödel's Theorem," *Philosophia Mathematica*, IX (2001): 257–79.

0022-362X/04/0104/197–210

> And that surely cannot be! And if it is proved, then it is proved that it is not provable. Thus it can only be true, but unprovable."
>
> Just as we can ask, "'Provable' in what system?," so we must also ask, "'True' in what system?" "True in Russell's system" means, as was said, proved in Russell's system, and "false" in Russell's system means the opposite has been proved in Russell's system.—Now, what does your "suppose it is false" mean? *In the Russell sense* it means, "suppose the opposite has been proved in Russell's system"; *if that is your assumption* you will now presumably give up the interpretation that it is unprovable. And by "this interpretation" I understand the translation into this English sentence.—If you assume that the proposition is provable in Russell's system, that means it is true *in the Russell sense*, and the interpretation "P is not provable" again has to be given up. If you assume that the proposition is true in the Russell sense, *the same* thing follows. Further: if the proposition is supposed to be false in some other than the Russell sense, then it does not contradict this for it to be proved in Russell's system. (What is called "losing" in chess may constitute winning in another game.)[3]

In this passage, we find Wittgenstein criticizing a relatively common interpretation of Gödel's first theorem: that the theorem shows—or helps to show—that there are true but unprovable sentences of ordinary number theory.[4] Wittgenstein objects to this interpretation, partially because he is skeptical concerning the notion of "Truth" in play here ("'True' in what system?"), and partially because he is opposed in principle to the derivation of "philosophical" claims from "mathematical" arguments.[5]

In their analysis of this passage, Floyd and Putnam focus on the following key claim (paraphrased from the second paragraph of Wittgenstein's discussion):

> (KC) If one assumes that $\neg P$ is provable, then one should give up the "translation" of P by the English sentence 'P is not provable'.

In what follows, I start by sketching the basic mathematics behind this claim. I then examine Floyd and Putnam's argument for the claim and what they take to follow from it. Finally, in sections II and III, I explain how and why their analysis goes wrong.

---

[3] *Remarks on the Foundations of Mathematics*, G.E. von Wright, R. Rees and G.E.M. Anscombe, eds., Anscombe, trans. (Cambridge: MIT, 1956), I, Appendix III, §8.

[4] Wittgenstein focuses on number theory as formulated in "Russell's system"—that is, the system of *Principia Mathematica*. There is, however, nothing in his argument that depends on this particular choice of background logic. For expository convenience, I will recast the argument in terms of ordinary, first-order Peano Arithmetic. Later—in section IV—I discuss the possible philosophical significance of formulating the argument in "Russell's system."

[5] See, for example, *Remarks*, VII §19. See also *Philosophical Investigations*, Anscombe, trans. (New York: Macmillan, 1953), §124.

I begin with the relevant mathematics. In his original paper on incompleteness,[6] Gödel defines two formulas in the language of formal number theory. These formulas—which I will call "PROOF(x, y)" and "SUBST(x, y, z)"—have the following nice property:[7]

> NUMERALWISE REPRESENTABILITY: Let m be the code of a sentence, ϕ, and let n be a natural number:
>
> (1) If n is the code of a proof of ϕ, then $\mathrm{PA} \vdash \mathrm{PROOF}[\hat{n}, \hat{m}]$.
> (2) If n is not the code of a proof of ϕ, then $\mathrm{PA} \vdash \neg\mathrm{PROOF}[\hat{n}, \hat{m}]$.
>
> Similarly, let $\phi(\upsilon_0)$ be a formula with one free variable, let p be the code of ϕ, and let n and m be natural numbers:
>
> (1′) If m is the code of $\phi(\hat{n})$, then $\mathrm{PA} \vdash \mathrm{SUBST}[\hat{p}, \hat{n}, \hat{m}]$.
> (2′) If m is not the code of $\phi(\hat{n})$, then $\mathrm{PA} \vdash \neg\mathrm{SUBST}[\hat{p}, \hat{n}, \hat{m}]$.

Using these formulas, Gödel defines the sentence which Wittgentein later calls "P" (although Gödel himself calls it something different). Initially, he defines the formula:

$$\psi(\upsilon_0) = \exists y\, [\mathrm{SUBST}(\upsilon_0, \upsilon_0, y)\ \&\ \neg\exists z\, \mathrm{PROOF}(z, y)]$$

Letting e_0 be the code for $\psi(\upsilon_0)$, Gödel sets:

$$P = \psi(\hat{e}_0) = \exists y\, [\mathrm{SUBST}(\hat{e}_0, \hat{e}_0, y)\ \&\ \neg\exists z\, \mathrm{PROOF}(z, y)]$$

This, then, gives the basic definition underlying Wittgenstein's (and later Floyd and Putnam's) discussion. There are two things we should notice about this definition.

First, the property of numeralwise representability explains why someone might think that P should be interpreted as saying "P is unprovable." On the assumption that PA is sound—that is, that $\mathbb{N} \models \mathrm{PA}$—the numeralwise representability of PROOF(x, y) and SUBST(x, y, z) entails the following, somewhat more semantic, property:[8]

---

[6] Gödel, "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems," in *Collected Works* (New York: Oxford, 1986), pp. 145–99.

[7] To unpack the terminology/notation here, I note that "coding" is simply a way of associating natural numbers to syntactic objects like formulas and proofs; done properly, it allows syntactic properties like "being well-formed" or "being a valid proof" to be treated as number-theoretic properties of the associated codes. Similarly, the notation $\hat{n}$ simply represents the expression for the number n in the language of formal number theory: specifically, $0 + 1 + \ldots + 1$, with n 1's.

[8] It is important to note that these facts about arithmetic expressibility can be proved directly, without passing through numeralwise representability. (Gödel himself approaches arithmetic expressibility this way in his 1931 paper.) Nonetheless, once we have numeralwise representability, it provides a quick route to arithmetic expressibility.

> ARITHMETIC EXPRESSIBILITY: For any two numbers n and m, $\mathbb{N} \models \mathrm{PROOF}(\hat{n}, \hat{m})$ if and only if m is the code of a sentence, ϕ, and n is the code of a proof of ϕ.
>
> Similarly, for any three numbers n, m, and r, $\mathbb{N} \models \mathrm{SUBST}(\hat{n}, \hat{m}, \hat{r})$ if and only if n is the code of a formula, $\phi(\upsilon_0)$, and r is the code of the sentence $\phi(\hat{m})$.

Using arithmetic expressibility, therefore, we can make perfect sense of the claims that $\neg\exists z\, \mathrm{PROOF}(z, y)$ "says that" the sentence coded by y is unprovable and that $\mathrm{SUBST}(\hat{e}_0, \hat{e}_0, y)$ "says that" y is the code of P. Together, these two claims provide the intuitive basis for interpreting P as "P is unprovable."

Second, our definition explains why someone might think that P is a true but unprovable sentence of number theory. If we filter P through arithmetic expressibility, we get the following result:

$\mathbb{N} \models P$ if and only if there is an n, such that $\mathbb{N} \models \mathrm{SUBST}(\hat{e}_0, \hat{e}_0, \hat{n})$ and $\mathbb{N} \models \neg\exists z\, \mathrm{PROOF}(z, \hat{n})$.

if and only if there is an n, such that n is the code of P and $\mathbb{N} \models \neg\exists z\, \mathrm{PROOF}(z, \hat{n})$.

if and only if there is no m, such that m is the code of a proof of P.

if and only if $\mathrm{PA} \nvdash P$.

This equivalence leaves us with two options: *either* $\mathbb{N} \models P$ and $\mathrm{PA} \nvdash P$ *or* $\mathbb{N} \not\models P$ and $\mathrm{PA} \vdash P$. Notice, however, that the second of these two options contradicts the soundness of PA (since it trivially entails that $\mathbb{N} \not\models \mathrm{PA}$). Hence, we are forced to accept the first option. And this first option looks an awful lot like the claim that P is true but unprovable.

This, therefore, gives us a sketch of the claims Floyd and Putnam want to argue against (that P says "P is unprovable" and that P itself is "true but unprovable"). To understand their objections to these claims, we can begin by looking at their argument for KC.[9] So suppose, just for the sake of argument, that PA is consistent but that we have discovered that $\mathrm{PA} \vdash \neg P$. The first thing to notice is that this entails that $\mathbb{N} \not\models \mathrm{PA}$ (that is, since $\mathbb{N} \models \mathrm{PA} \Rightarrow \mathbb{N} \models \neg P \Rightarrow \mathbb{N} \not\models P \Rightarrow \mathrm{PA} \vdash P \Rightarrow$ PA is inconsistent). It follows, therefore, that PA is satisfied *only* by nonstandard models of number theory—that is, by models which are not isomorphic to the natural numbers. To use the technical jargon, it follows that PA is ω-*inconsistent.*[10]

---

[9] The argument occurs on pp. 625–27.

Consider, then, a specific model of PA. Call this model $\mathbb{M}$, and call the code of P, p_0. Then, since PA $\vdash \neg P$, we know that $\mathbb{M} \models \exists x\, \text{PROOF}(x, \hat{p}_0)$. Therefore, there is some element $m \in \mathbb{M}$ such that $\mathbb{M} \models \text{PROOF}[m, \hat{p}_0]$. But, by numeralwise representability, $\mathbb{M} \models \neg\text{PROOF}(\hat{n}, \hat{p}_0)$ for each natural number n. Hence, the relevant m is not one of the ordinary natural numbers—it is, of necessity, one of the "nonstandard" elements of $\mathbb{M}$.[11] Further, and this is the second thing to notice here, there is no interesting sense in which this nonstandard m "codes up" a proof of P—or of any other formula, for that matter. Given this, we have no reason to think that the formula PROOF(x, y), *as it gets interpreted by the model* $\mathbb{M}$, still captures the notion "y is the code of a sentence and x is the code of a proof of that sentence."

In this situation, therefore, Floyd and Putnam argue that it is hard to justify interpreting P as meaning "P is not provable." After all, the intuitive motivation for interpreting P this way involved interpreting P on the natural numbers—that is, interpreting "$+$" as plus, "$\times$" as times, and letting "$\exists x$" range over the natural numbers. But now it turns out that this interpretation is incompatible with PA itself (since PA has *only* nonstandard models). Similarly, the interpretation of P as "P is not provable" owed something to the interpretation of PROOF(x, y) suggested by arithmetic expressibility. But, as we have just seen, this interpretation of PROOF(x, y) breaks down when we interpret PROOF(x, y) on nonstandard models. Hence, once we make the assumption that PA $\vdash \neg P$—and thus that *there are no standard models for* PA—we have every reason to give up our initial interpretations of both PROOF(x, y) and of P.[12]

This, then, is the core of Floyd and Putnam's argument for KC. What a given formula "expresses" depends on the model at which we interpret it. If we interpret PROOF(x, y) and P on the natural numbers, then it is plausible to think that they express the facts that "x codes a proof of y" and that "P is unprovable." If we interpret them on nonstandard models, then it is no longer plausible to think that they express such things. Therefore, once we assume that PA is ω-inconsistent—and hence that *all* "admissible interpretations" take place on nonstandard models—we are left with *no* good reasons for interpreting P as "P is not provable."

---

[10] Formally, ω-inconsistency is a bit stronger than the claim that a system has only "nonstandard" models to say that PA is ω-inconsistent means that there is some particular formula, $\phi(x)$, such that for each n, PA $\vdash \neg\phi(\hat{n})$, but it is also the case that PA $\vdash \exists x\, \phi(x)$. In the specific case where PA $\vdash \neg P$, we can actually identify the relevant $\phi(x)$: for each n, PA $\vdash \neg\text{PROOF}(\hat{n}, \hat{p}_0)$, but PA $\vdash \exists x\, \text{PROOF}(x, \hat{p}_0)$. So, in this case, it is the very formula used in the construction of P that serves to witness the ω-inconsistency of PA.

[11] Cf. the discussion of ω-inconsistency in the preceding footnote.

[12] In general, there may also be problems concerning the interpretation of SUBST(x, y, z). In particular, there will be nonstandard elements $m_1, m_2, m_3 \in \mathbb{M}$ such that $\mathbb{M} \models \text{SUBST}(m_1, m_2, m_3)$, even though it is clear that m_2 does not code an ordinary natural number and that neither m_1 nor m_3 code formulas. Fortunately, this does not have to affect the interpretation of P, since we *can* show that $\mathbb{M}$ interprets SUBST$(\hat{e}_0, \hat{e}_0, z)$ correctly (that is, $\mathbb{M} \models \text{SUBST}(\hat{e}_0, \hat{e}_0, m) \Leftrightarrow \mathbb{M} \models m = \hat{p}_0$).

So much for KC itself. With KC in hand, I turn to the consequences Floyd and Putnam draw from KC. To begin, KC shows that the *formal structure* of P does not force us to interpret P as "P is not provable" (since this is an interpretation we would *give up* under certain circumstances—for example, if we discovered that PA $\vdash \neg P$). Nor, Floyd and Putnam argue, do the other details of Gödel's *mathematics* force us to interpret P this way. As formulated in his original paper, Gödel's proof is purely syntactic (using numeralwise representability to show that if PA is ω-consistent, then PA $\nvdash P$).[13] Hence, as far as the *mathematics* goes, we can dispense with *interpretation* altogether.[14] Given all this, it is unclear where the claim that P should be interpreted as meaning "P is unprovable" is supposed to *come from.* At best, this seems to be a "metaphysical claim" (632) which gets grafted onto Gödel's mathematics; it is not, in any interesting sense, something which *follows from* that mathematics.

At the end of the day, Floyd and Putnam take themselves to have vindicated Wittgenstein's skepticism about the common insistence that Gödel's incompleteness theorem shows that there are true but unprovable sentences of ordinary number theory. To make this claim plausible, we seem to need an interpretation of P which makes P *say* "P is not provable." But KC—along with an analysis of the actual *mathematics* of Gödel's proof—undercuts the idea that Gödel's work provides such an interpretation. As a result, Floyd and Putnam conclude that the common insistence that P is "true but unprovable" is more a "metaphysical claim" than a "mathematical result" (632). And, while this does not entail that the claim is *false*, it does show "how little sense we have succeeded in giving it" (632).

---

[13] As Mic Detlefsen has pointed out to me, the claim that Gödel's proof is "purely syntactic" needs to be treated with some care. On the one hand, Gödel's proof does not require specifying a full, formal interpretation of the symbols of his language. (Gödel himself emphasizes this fact on pp. 171, 177, and 181.) Nor does it involve a notion of *truth* for that language. So, it is not semantic in the way that a model-theoretic argument is semantic (or one which made use of arithmetic expressibility).

On the other hand, the argument does involve a systematic *association* between natural numbers and terms in our language (for example, between n and $\hat{n}$). This association is crucial for proving the numeralwise representability results on which the overall proof depends. So, even Gödel's "syntactic proof" involves general correlations between natural numbers and the formal expressions which, in Gödel's own terms, "denote" those numbers.

[14] At best, Floyd and Putnam suggest, Gödel's paper gives rise to a proof-theoretic conception of truth—one under which "P is true" *means* PA $\vdash P$ and "P is false" *means* PA $\vdash \neg P$. Clearly, however, this is an interpretation which leaves no room for "true but unprovable" sentences of arithmetic.

Now, for the reasons mentioned in the last footnote, I find the claim that Gödel's mathematics does not involve "interpretation" somewhat problematic (though there is clearly *a* sense in which it is true). For the sake of argument, though, I will avoid these complexities by simply granting the claim that Gödel's proof is "purely syntactic" and that it involves no "interpretation."

II. ON KC

In this section, I step back to examine Floyd and Putnam's argument for KC in more detail. (In section III, I will look at their more general conclusions concerning the incompleteness theorem.) To see what is wrong with Floyd and Putnam's argument for KC, we need to back up a bit and recall the hypothetical situation KC envisages. Initially, we come to Gödel's theorem with an interpretation of the language of formal arithmetic on the natural numbers—an interpretation which reads "$+$" as plus, "$\times$" as times, and which lets "$\exists x$" range over the natural numbers.[15] Given this interpretation, we proceed to assume two things. First, we assume that our background axiomatization of arithmetic is *sound*—that is, that all of the axioms of PA come out *true* on our interpretation. (If we use standard model-theoretic machinery, this amounts to the claim that $\mathbb{N} \models$ PA.)[16] Second, on the basis of an argument like that given on page 200, we assume that we can interpret P as "P is not provable."

Now at this point, KC suggests the following hypothetical: suppose PA $\vdash \neg P$. As we saw in section I, this entails that $\mathbb{N} \not\models PA$. Hence, we face a choice. On the one hand, we could modify our background interpretation of arithmetic, giving up $\mathbb{N}$ as an appropriate model for our language and limiting ourselves to those (nonstandard) models which happen to satisfy PA. If we make this choice, then we will be forced to give up the interpretation of P as "P is not provable" (for the reasons highlighted by Floyd and Putnam). On the other hand, we could keep our background *interpretation* of arithmetic language and give up the assumption that PA provides a satisfactory *axiomatization* of arithmetic.

---

[15] After Tarski, this interpretation would probably be fleshed out using model-theoretic machinery. But this is not necessary. Gödel himself sketches a non-Tarskian version of the interpretation, and it is this non-Tarskian interpretation which Gödel uses in formulating the notion of arithmetic expressibility (and in proving things about that notion); see pp. 181–91.

[16] For the remainder of the paper, I will assume that we *have* used standard model-theoretic machinery in fleshing out our interpretation. Hence, I will use $\mathbb{N} \models$ PA to express the soundness of our axioms.

It is clear from their paper that Floyd and Putnam think we should take the first option (as witnessed, for instance, by their insistence that only models which satisfy PA should count as "admissible interpretations" for our language).[17] But it is equally clear that they provide no real *argument* for preferring this option. Instead, they simply ignore the possibility of *keeping* $\mathbb{N}$ as the canonical interpretation for our language while *abandoning* (or, at the very least, *modifying*) the ω-inconsistent axiomatization that stands in conflict with this interpretation.[18]

So far, this is simply an objection to Floyd and Putnam's *argument* (pointing out a lacuna which could, in principle, be filled in). But there is a deeper problem here: Floyd and Putnam's assumptions about the (hypothetical) response of the mathematical community to the discovery that $\mathrm{PA} \vdash \neg P$ are almost certainly false. Although such a discovery would cause a great deal of consternation, I think the vast majority of mathematicians would look for ways of revising PA in order to block the proof in question—that is, would try to isolate the specific axioms of PA that are essential to the proof and to eliminate some of them from our axiomatization.[19] There are three points to make about this.

First, it is almost *unimaginable* that mathematicians would adopt recognizably nonstandard models of arithmetic as canonical for interpreting the language of number theory. Neither would they accept a provably ω-inconsistent axiomatization of arithmetic as adequate (since it would not, after all, describe the natural numbers!). Given this, the straightforward claim that mathematicians would reject the interpretation of P as "P is not provable" *because* they accept nonstandard models as the basis for interpreting arithmetic is surely mistaken

---

[17] The insistence that we limit ourselves to models of PA when we interpret arithmetic runs rather deep in Floyd and Putnam's paper. At one point they even suggest that, were we to find PA *inconsistent*, we should conclude that there are *no* admissible interpretations of arithmetic (see 626).

[18] This is not, perhaps, so surprising in Putnam's case, as a similar insistence on the priority of axioms over interpretations lies at the heart of his so-called "model-theoretic argument" against realism. See sections III–IV of Timothy Bays, "On Putnam and His Models," this JOURNAL, XCVIII, 7 (July 2001): 331–50.

[19] My guess is that mathematicians would initially focus on the uses of induction in the proof. The hope would be that some well-motivated restriction of the induction scheme would enable us both to restore ω-consistency and to understand *why* our initial scheme went wrong (for example, perhaps we allowed induction on some subtly-paradoxical predicate/formula). This seems *far* more likely than the abandonment of $\mathbb{N}$ which Floyd and Putnam urge upon us.

(though it is equally surely what Floyd and Putnam's argument requires).

Second, although mathematicians would not abandon the interpretation of *P* as "*P* is not provable" for the reasons Floyd and Putnam suggest, there is one reason they might abandon it. Suppose it turns out that the specific mathematics used in proving arithmetic expressibility is somehow implicated by the discovery that $\mathrm{PA} \vdash \neg P$. (Say, because there is some $\phi \in \mathrm{PA}$ such that (1) ϕ is crucial to the proof of $\neg P$ and (2) there are numbers *n* and *m* such that *m* codes a formula, and *n* codes a proof of that formula, but $\mathrm{PA}\backslash\{\phi\} \nvdash \mathrm{PROOF}(\hat{n}, \hat{m})$.)[20] In such a case, the modifications to PA which are needed to *avoid* the acceptance of nonstandard models might undercut various arithmetic expressibility results. In doing so, they might also undercut the traditional interpretation of *P*.

This, then, looks like a circumstance where something like KC might really be true. Still, there are some obvious worries. First, the supposition in the last paragraph is extremely implausible. The numeralwise representability results on page 199 can be proved in *very* weak fragments of arithmetic, and it is hard to imagine these fragments being problematic.[21] Second, any discovery of problems in weak arithmetic—say, in Q_0—would require such deep revisions of present mathematics that it is virtually impossible to adjudicate questions concerning "what we would/should do" in such circumstances. In particular, neither I nor Floyd and Putnam are in a position to intelligently evaluate KC under this kind of hypothetical. Finally, whatever we would do under this hypothetical, we would do it for reasons other than those Floyd and Putnam suggest. If, for instance, we would abandon the interpretation of *P* as "*P* is not provable," then we would do so as part of a larger revision of PA aimed at *keeping* the standard model of arithmetic; we would not accept nonstandard models in order to keep the ω-inconsistent PA.

This brings me to my third point: nothing I have just said implies that mathematicians would *lose interest* in PA and its models (that is, under the circumstances envisioned by KC). Just as logicians now study nonstandard models of arithmetic (while acknowledging that these models *are* nonstandard), so mathematicians would continue to think about nonstandard models after the discovery that PA $\vdash \neg P$ (indeed, they would probably study them a lot more than they do now). Nevertheless, the majority of mathematical work—and the work that would be universally recognized as *arithmetic*—would continue to involve the study of $\mathbb{N}$: it would involve interpreting the language of arithmetic *on* $\mathbb{N}$, and it would look for axioms satisfied *by* $\mathbb{N}$.[22]

---

[20] Of course, this particular result would simply block the proof of arithmetic expressibility *via* numeralwise representability. The crucial case—and the case we are really interested in—is where the elimination of ϕ would undercut *every* proof of arithmetical expressibility (perhaps by undercutting the recursive definition of satisfaction). Nevertheless, the case above illustrates the *kind* of problem at issue here.

[21] So, for instance, numeralwise representability holds for systems as weak as Robinson's Q_0 (a fragment of arithmetic which involves *no* induction); it is very hard to imagine Q_0 being ω-inconsistent.

These, then, are some reasons for thinking that Floyd and Putnam's argument for KC is flawed and that KC itself is mistaken. Before moving on, I think it is useful to digress a moment and examine a claim similar to KC outside the mathematical context. Suppose we have a formal axiomatization of some part of physics—call it T and suppose that it is formulated in the language $\mathcal{L}$. Next, suppose we discover that T proves things that conflict with the basic physical phenomena T was supposed to describe (for example, T makes a series of glaringly false predictions). Then, just as in Floyd and Putnam's case, we face a choice: we can abandon T and start looking for new axioms that better describe the phenomena we are interested in, or we can abandon the physical phenomena and start studying arbitrary models of T (numerical models, perhaps).

Here, I think it is obvious that the physics community would take the first option. T was interesting only (or, at least, primarily) because it seemed to describe a specific class of physical phenomena (that is, described it via the standard interpretation of $\mathcal{L}$). Once T goes wrong about such phenomena, then T has to be modified. In the scientific case, therefore, preserving the original interpretation of our language turns out to be far more important than preserving our original axiomatization. I see no reason to think that things are different for the mathematical case; hence, I see no reason to accept Floyd and Putnam's argument for KC. At the very least, the structural similarities between their argument and the (obviously faulty) physical analogue should lead us to be highly suspicious of the former.

III. ON GÖDEL'S THEOREM

In this section, I turn from Floyd and Putnam's defense of KC to make some brief comments concerning the conclusions they draw from this claim. First, I concede a point: there is nothing in the formal structure of *P*—that is, in *P*'s very syntax—which forces us to interpret *P* as "*P* is not provable." Nor, as Floyd and Putnam notice, does Gödel's original proof require such an interpretation. So, if these were the only ways of providing a mathematically significant interpretation of *P*, then Floyd and Putnam would be right in challenging the claim that *P* means "*P* is not provable."

---

[22] A qualification is in order here. A lot of current work in arithmetic focuses on the purely algebraic properties of the natural numbers (and associated structures like $\mathbb{Z}$ and $\mathbb{Q}$). To the extent that nonstandard models of $\mathbb{N}$ (and the associated nonstandard versions of $\mathbb{Z}$ and $\mathbb{Q}$) continue to satisfy the relevant algebraic axioms, people working on algebraic number theory may not care about nonstandardness. Nevertheless, no matter how interesting these nonstandard models are (to algebrists and logicians), the canonical core of arithmetic will continue to be the study of $\mathbb{N}$, and the axioms for arithmetic will have to be axioms satisfied by $\mathbb{N}$.

Fortunately, these are *not* the only respectable ways of interpreting *P*. The natural interpretation of the language of arithmetic—the interpretation under which "$+$" means plus, "$\times$" means times, and "$\exists x$" ranges over the natural numbers—can be made perfectly rigorous. At present, it would be most natural to do this using model-theoretic machinery or Tarski's original apparatus of language, metalanguage, meta-metalanguage, and so forth. But even without this machinery, we can formulate the interpretation in a mathematically perspicuous manner. Indeed, Gödel himself uses this interpretation in his 1931 paper, both to formulate the notion of arithmetic expressibility and to prove theorems involving this notion.[23]

Further, this interpretation has closer ties to the incompleteness theorem than Floyd and Putnam's argument would lead one to believe. I have already noted that Gödel *uses* the interpretation in several parts of his paper, and he seems to *presuppose* it in many of his informal asides.[24] More substantially, the interpretation helps to explain why Gödel's theorem is *interesting* in the first place. The theorem is not interesting because it shows that a randomly selected axiom system happens to be incomplete: there are *lots* of incomplete axiom systems, and most of them are rather boring.[25] Instead, the theorem shows that a standard axiomatization of *arithmetic*—of the theory of $\mathbb{N}$—is incomplete. More significantly, the theorem does not just show that *this particular* axiomatization is incomplete. Rather, it shows that arithmetic is *intrinsically* incomplete: no recursive extension of PA (or even of Q_0) provides a complete axiomatization of arithmetic.[26] It is this intrinsic incompleteness of *arithmetic* which makes Gödel's first theorem so interesting.

Finally, and most importantly, although it is certainly true that Gödel's original proof of incompleteness was syntactic, there is an alternate, semantic proof which makes essential use of the interpretation at issue here. This proof starts with the arithmetical expressibility results from page 200 and then uses the argument which immediately follows those results (again, page 200) to show that P is a true but unprovable sentence (and, hence, that P is undecidable from PA).[27] So, even though Gödel did not use this interpretation in his official proof of incompleteness, the interpretation is still very closely related to the underlying *mathematics* of Gödel's paper.[28]

---

[23] See Gödel, p. 181 for the definition of arithmetic expressibility. See pp. 183–87 for some theorems which use this notion.

[24] See Gödel, pp. 149–51.

[25] For example, the systems $\emptyset$, $\{P(c)\}$, and $\{c \neq d\}$ are all trivially, and uninterestingly, incomplete.

[26] So, there is no way to "fix" the incompleteness of PA by adding a few new axioms.

When all is said and done, then, I think that Floyd and Putnam are mistaken in the conclusions they draw from KC. There is a perfectly good—and a perfectly *mathematically* respectable—interpretation of the language of arithmetic under which P expresses the fact that P is not provable. On this interpretation, Gödel's theorem really does show that P is "true but not provable." Further, this interpretation is in no way foreign to Gödel's work. Gödel uses the interpretation in the paper where he proves his incompleteness theorem; the interpretation helps to explain the *significance* of that theorem; and the interpretation can be used—as Gödel well knew—to provide a rigorous proof of that theorem. All of this is plain, simple mathematics; there are no "metaphysical claims" anywhere on the horizon.

IV. "RUSSELL'S SYSTEM" AND FOUNDATIONS

Before concluding, I want make two remarks concerning an issue I bypassed near the beginning of this paper: the fact that I have formulated my arguments in terms of PA, while Wittgenstein, Floyd and Putnam formulated theirs in terms of "Russell's system." Now on one level, any reversion back to Russell's system would only make my argument stronger. PA is a far more widely accepted formal system than Russell's system is (or ever was). Hence, just as it is clear that we would modify PA to deal with a discovery that PA is ω-inconsistent (as argued in section II), it is *even more clear* that we would modify Russell's system to deal with an ω-inconsistency in that context.[29] As a result, the arguments I gave in sections II and III would be even stronger if KC were reformulated in terms of Russell's system. So, there is nothing "slippery" in moving the arguments to PA for reasons of perspicuousness.

---

[27] It is worth noting that this semantic proof amounts to a formalization of the informal proof given in the first paragraph of Wittgenstein's remarks (see pages 197–98). Wittgenstein, however, goes on to criticize this proof in his second paragraph. Surprisingly, it is just this criticism which Floyd and Putnam mean to be defending. See Steiner for more on Wittgenstein's rejection of the semantic proof.

[28] Two historical comments are in order. First, Gödel himself was clearly aware of the semantic proof of incompleteness: he sketches it at the beginning of his 1931 paper (pp. 149–51), and he outlines it again in a series of lectures given at Institute for Advanced Study in 1934. Second, Gödel had several reasons for avoiding this semantic proof in his 1931 paper. For one thing, the syntactic proof provides crucial "ingredients" for proving the second incompleteness theorem (while the semantic proof does not). For another, Gödel intended his theorem to have relevance to David Hilbert's program, but only the syntactic proof of the theorem could have such relevance. Hence, Gödel had a number of purely mathematical reasons for using a syntactic proof in 1931. Given this, Gödel's use of the syntactic proof should not be interpreted as a lack of *awareness* of the semantic proof (and still less as a *rejection* of that proof).

[29] Two technical issues deserve comment here. First, the notion of ω-inconsistency depends on the version of Russell's system with which we are working. Gödel himself

On another level, though, there may seem to be a problem here. Russell wanted his system to provide a *foundation* for mathematics: to provide the framework in which other parts of mathematics are formulated, and to set the standards for mathematical rigor. Given this, it might seem odd to think that we can "step outside" this framework to discuss its semantics (and then *reject* the framework if it does not live up to our expectations!). To put the point in Floyd and Putnam's terms: "to confess that this is what one has to do would be to abandon the claim for the *foundational* status of a system such as *Principia Mathematica* entirely" (630). On this view, then, it may seem that my entire argument in sections II and III depends on adopting the wrong *attitude* toward our background axiomatization (on treating it as "just one piece of mathematics among others").[30]

Although this line may initially be tempting, it should be rejected for two reasons. First, there *cannot* be a problem simply with stepping back to study the metatheory of our axiomatization. For, if that were the problem, then it would indict the study of *proof theory* as much as the study of semantics (and, hence, indict the syntactic argument in Gödel's original paper). Any version of Gödel's theorem will involve some amount of "stepping back" from our foundational system (that

---

worked with a version of Russell's logic which was "superimposed" on ordinary Peano Arithmetic: in particular, Gödel's formulation proved that "every object is a number." For this formulation of Russell's system, the definition of ω-inconsistency given in footnote 10 works just fine.

If we work with the full system of PM—or, for that matter, with a system like ZFC—then our system will prove that there are objects other than numbers (and we will have to use a predicate or formula to pick out the particular objects which we want to count as "natural numbers"). For such systems, we say that T is ω-inconsistent if there is a formula, $\phi(x)$, such that for each n, $T \vdash \neg\phi(\hat{n})$, but it is also the case that $T \vdash \exists x\,[\mathrm{NUMBER}(x)\ \&\ \phi(x)]$. This is a/the standard way of thinking about ω-inconsistency for systems which talk about more than number theory. (Note that if we simply used the definition from footnote 10, then systems like T would be *trivially* ω-inconsistent.)

Second, the discovery that PM was ω-inconsistent would leave us with several *different* options. We could modify the underlying *logic* of PM; we could change the way we formulate arithmetic *in* PM; or we could leave everything as it is and cease to regard our system as a formulation *of arithmetic* (see the comments in the main text vis-à-vis this last option). The only thing we cannot do, I contend, is to leave things as they are *and* insist that we still have a formulation of *arithmetic*.

[30] For what it is worth, Gödel himself seems to share my "bad attitude" toward Russell's system. For Gödel, Russell's system was simply one of several "related systems" to which his results applied, and he felt free to step back and discuss both the syntax and the semantics of each of these systems.

is why it is a result in *metamathematics*), and I see no principled reasons for permitting us to step back to make syntactic generalizations, while forbidding us to step back to make semantic generalizations.[31] Therefore, if the foundationalist worry is to be cogent, it has to focus on the claim that metatheoretic analysis can make us *give up* a foundational system (and not on the very idea of metatheoretic analysis).

Second, there are clear cases where we *can* be forced to give up a foundational system (for example, when it proves a contradiction).[32] And even when we do not give up a foundational system—because it is too clear, precise and elegant to give up—we can still reject it *as a foundation for a specific discipline.* If a "foundation for arithmetic" winds up proving that $2 + 2 = 5$, then it is not a foundation for arithmetic (though it may well be foundation for something else). Similarly, if a "foundation for analysis" proves that the derivative of $x^3 + 1$ is x^{27} then it is not a foundation for analysis (though, again, it may be a foundation for something else).[33] It is just a *crazy* view of foundational research which holds that, once a foundational system has been proposed, we can no longer legitimately discuss whether that system is *effective* at founding the subjects we wanted it to found in the first place.

In the end, then, I do not see any problems with either my move from Russell's system to PA or my willingness to consider modifications of "foundational" axiom systems. In particular, I do not see any "foundationalist" reasons for changing the conclusions I reached in sections II and III. There is a perfectly good—and, indeed, a perfectly canonical—interpretation of arithmetic under which Wittgenstein's *P* really does say "*P* is not provable." Given this interpretation, Gödel's theorem helps to show that there are "true but unprovable" sentences of ordinary number theory. Nothing in Wittgenstein's remarks—or in Floyd and Putnam's analysis of those remarks—should lead us to think otherwise.

TIMOTHY BAYS

University of Notre Dame

---

[31] As a point *tu quo,* I would note that Floyd and Putnam themselves engage in purely semantic reasoning about Russell's system. So, for instance, they argue that it would be acceptable to define "truth in PM" as "holding in all models of PM" (631). Similarly, they are willing to prove that, if PM $\vdash \neg P$, then *all models of PM* are nonstandard (625). Given this, it would be unreasonable for them to object *in principle* to the semantic analysis of foundational systems.

[32] The important point, here, is that we can recognize from the *outside* that contradictions are unacceptable. We do not just accept contradictions because our "foundational" system tells us to. Hence, we certainly have *some* ability to step back and engage in critical reflection on (purported) foundations.

[33] And, I would argue, if a "foundation for arithmetic" winds up being ω-inconsistent, then it is not really a foundation for arithmetic (whatever other virtues it might happen to have).

BOOK REVIEWS

Reasonably Vicious. CANDACE VOGLER. Cambridge: Harvard University Press, 2002. viii + 295 p. Cloth $47.00.*

Candace Vogler's insightful and provocative book takes on several projects that unfold one from another, as if new philosophical aims were being discovered along the way. Its initial aim is to explain why it is that instrumentalist theories of practical reasoning have such grip on us, such that there seems to be a recurring need to attempt to refute them. In explaining that, Vogler distinguishes between instrumentalism as such, which includes a "bifurcationist" moral psychology that radically differentiates the cognitive from the volitional, and "calculative" theories of practical reasoning, which do not, but instead concentrate on the end-means form of "practical justification." Her explanation of instrumentalism's grip on us is, roughly, that instrumentalism is easily confused with the view that all practical justification is calculative. The second task, then, becomes shoring up the plausibility of the latter thought. The third layer of the project concerns the particular way that she sets out to shore this up, namely by resurrecting Aquinas's division of worldly goods into the useful, the pleasant, and the fitting. Five of the book's eight chapters are devoted to elaborating and expounding this division. Finally, as an apparently independent aim, and seemingly as an afterthought, the book arrives at its title theme, addressing the connection between calculative practical reasoning and morality. Vogler argues that an amoralist does not necessarily suffer from any calculative practical irrationality.

Vogler's initial diagnostic chapter usefully distinguishes among (a) the metaphysics and epistemology of value, (b) the moral psychology of action, (c) the structure of practical reasoning, and (d) the structure of practical justification (12). Her principal point in making these distinctions is to suggest that the opponents of so-called "instrumentalism" have been sloppy about them and have supposed that a thesis in one of these categories (such as the assertion of a Humean, bifurcationist moral psychology) necessarily combines with a thesis in another (such as the assertion that the structure of practical justifi-

---

* This review has benefited immeasurably from extensive discussion of Vogler's book with William Haines, Samuel Kerstein, Mark C. Murphy, Nancy Sherman, and Karen Stohr.

0022-362X/04/0104/211–15

cation is necessarily "calculative," or end-means). The first two categories are dropped from consideration early on, and, as the book proceeds, it becomes increasingly clear that Vogler is devoting all of her attention to the last, the "structure of practical justification." How are the final two categories distinguished?

The distinction Vogler is drawing between the structure of practical reasoning and the structure of practical justification becomes fully clear only in chapter 7. There, she addresses those unnamed philosophers who have argued that sound practical reasoning can settle upon new final ends that are not means (causal or constitutive) to ends with which the deliberator started (157). In response, Vogler argues that even assuming this is possible, still, deliberation itself is an action or activity, and as such proceeds with an end in view, namely *figuring out what to do* (166). Here, she might have quoted Kant: "every action has its end."[1] Accordingly, all deliberation or practical reasoning, no matter what its structure or content, is calculative in *form*, as it is directed to figuring out what to do (169).

What follows from this seemingly innocuous observation? Figuring out what Vogler takes to follow from it is complicated by the book's intermixture of diagnosis and more straightforward argument. For many of its pages, it seems she may simply be aiming to give the calculative view a run for its money in order to indicate how instrumentalist views could ever have had such a hold on us, whereas in other places she is clearly defending theses in her own voice. Among the theses she appears to be defending are a weaker and a stronger, each thrice repeated. The weaker thesis amounts to a denial of the kind of reasons externalism long ago defined by the late Bernard Williams:

> Noncalculative theorists insist that there can be reasons for an agent to *B* quite apart from her plans, projects, particular attachments, character, skills, habits, or further ends; the calculative theorist denies this (50, 150, 185).

Thus, at one level, Vogler's defense of the calculative theory of practical reasoning comes down to insisting that practical reasoning can generate reasons for action only by virtue of drawing on ends that antecedently have a foothold in the agent's motivational set. Unlike bifurcationist defenders of internalism, however, Vogler seems happy to let the truth of the calculative theory, so described, rest on the conceptual fact that, because practical deliberation is a kind of action, it necessarily begins with an end in view, namely that of figuring out what to do. If this is all that internalism, or the calculative view, entails, then there is indeed little reason to resist it.

---

[1] *The Metaphysics of Morals* Mary Gregor, trans. (New York: Cambridge, 1996), Ak. 385.

Vogler also provides a stronger statement of the calculative thesis about practical justification, however—one that she is also out to defend:

> Let *A* and *B* be actions of different types. One has reason to *A* only if one takes it that *A*-ing is a means to (or part of) attaining (or making it possible to attain) a further end, *B*-ing, and one wants to *B* (for no particular reason or because *B*-ing is pleasant, useful, or fitting) (48, 73, 149).

Here is a seemingly controversial version of the calculative theory—and also the link to Aquinas's tripartite division of goods. The unnamed philosophers who argue that sound practical reasoning can establish new ends would presumably want to contest this restriction on reasons for action—or would they?

Before we can answer this question, we must attempt to clear up some apparent anomalies in this stronger formulation of the calculative theory. We need to understand (i) why this thrice-repeated formulation would be structured so as apparently to demand a regress of practical reasons, (ii) how it can help make intelligible that one has any practical reasons, and (iii) why it requires that the agent believe there to be an end-means connection between *B* and *A*. The first issue arises because this is plainly put forward as a universal generalization across actions or action-types. Suppose that one has located a pair of actions, *A* and *B*, that one takes to be related as means to end, and that one wants to *B*. Then one has satisfied this necessary condition for having a reason to *A*. But does one have a reason to *B* just by virtue of wanting to (for no particular reason, say)? No, for the thesis again applies to *B*: there must be some other action *C* to which *B* is related as a (causal or constitutive) means if the agent is to have any reason to *B*. And so on *ad infinitum*. Vogler nowhere addresses this regress of reasons as a worry: that is the first puzzle with this stronger formulation. It is intimately connected with the second puzzle: How can this end-means relationship help make a reason for action intelligible if it is enough that *B* is something that the agent wants "for no particular reason"? If this answer is available, why is it not already available at the first layer, enabling the agent to evade this supposed necessary condition simply by saying "I felt like *A*-ing"? Finally, does it not seem crazy to hold that an agent has a reason for acting only if she "takes it" that a relevant end-means connection holds? Consider Williams's classic description of the agent who "believes that this stuff is gin, when it is in fact petrol. Has he reason, or a reason, to mix this stuff with tonic and drink it? On the one hand, it is just very odd to say that he has a reason to drink it.... On the other hand, if he does drink it, we not only have an explanation of his doing so (a reason why he did it), but we have such an explanation which is of the reason-for-action form."[2] Now, Vogler's stronger thesis states only a necessary condition for the existence of a practical reason, not a sufficient one, so she is not committed to saying that Williams's agent has a reason to drink the glass of gas. Conversely, however, she is committed to the claim that if, unbeknownst to the agent, the glass contains, not gin but an antidote to a poison that is about to kill her, still the agent has no reason to drink it unless she wants some gin. This seems equally odd.

Williams's description, however, provides the clue we need to resolve this puzzle as well as the other two. Although Vogler pointedly drops the theology to which each was committed, she follows G.E.M. Anscombe even more closely than she follows Aquinas. In particular, she finds Anscombe's concentration on actual actions, and the question why they were done, to be considerably more fruitful than the "deliberationist" question of how we reason about which actions to take. In Williams's original petrol/gin case, "if he does drink it," and we ask why he did, we *will* want to know what he took there to be in the glass. By the same token, if he did not take there to be any crucially needed antidote in the glass, then there will be no mystery why he refrained from drinking. Following Anscombe, Vogler is concentrating on the justification of actions actually done.

This Anscombian focus on the enterprise of explaining actions already done helps explain Vogler's differentiation, which may have seemed mysterious when presented above, between the structure of practical reasoning and the structure of practical justification. It also helps with the other two puzzles about her stronger formulation. The regress of reasons is not a worry, as there is no ambition to locate the ultimate normative grounds of action. Rather, to "justify" an action in the relevant sense, it appears, is simply to lay out the end-means form without which, as Aristotle and Anscombe have helped us see, no movement can be grasped as a case of action. That this is what Vogler means is confirmed by an elaborate thought experiment in chapter 6, in which she asks us to imagine what the practical world would look like if the "calculative order" were wholly drained out of it. Affirming that "the calculative form revealed by answers to Anscombe's 'Why?' question belongs to intentional action as such," she notes that if the calculative relation were removed from the world, actions would decompose into meaningless fragments (130f.). Thus, to the question "Why are you crossing the road?" you would have available only the answer, "because I felt like it" (or "for no particular reason"). But, she suggests, that is also the only answer one would have to the questions "Why are you stepping off the curb?" and "Why are you lifting your left foot?" Remove the calculative form, and intentional action as such collapses. Conversely, the "justification" with which she is concerned is provided so long as this calculative form, conceptually necessary to all intentional action as such, is restored. That is why a regress vainly in search of a stopping point to "Why?" questions is not a concern and why intelligibility is sufficiently provided, for Vogler's purposes, by indicating an action's calculative form.

---

[2] "Internal and External Reasons," in *Moral Luck* (New York: Cambridge, 1981), pp. 101–13, here see p. 102.

Accordingly, Vogler's stronger formulation, on this reading, is not nearly as controversial as it had appeared. The unnamed philosophers who defend the possibility of reasoning practically about final ends can accept even this stronger formulation, on the understanding that it is simply formulating a necessary condition on an agent's reasons in acting (as she actually is or has), not a necessary condition on the reasons that bear on an agent's deciding how to act. By the same token, it should be easy to accept her title claim: vicious actions can enjoy the "reasonableness" of the calculative form. In other words, evil action is possible.

With this interpretation taken as read, Vogler's explanation of why instrumentalism has had such a hold on us makes perfect sense: it has had a firm hold on us because all intentional action is calculative in form. As Kant said, "every action has its end." Setting aside, with Vogler, issues about the moral psychology of practical reasoning, what we can say is that instrumentalist theories of practical reasoning have retained a grip on us because of our persistent tendency to confuse the form of intentional action with the structure of practical reasoning. Calling this kind of calculative view a theory of practical "justification" is somewhat misleading; but as long as we pay sufficient attention to Vogler's frequent invocations of Anscombe and to her disinterest in "deliberative" accounts of practical reasoning, we will be able to remember that she is really talking about the reasoned structure of intentional action as such.

In thus attempting to clarify the core elements of Vogler's argument, I have not had space to describe her fascinating explorations of the pleasant, the useful, and the fitting. Philosophical readers are bound to find reading these chapters pleasant, useful, or a fitting use of their time.

HENRY S. RICHARDSON

Georgetown University

NEW BOOKS

SOMERVILLE, JAMES: *The Epistemological Significance of the Interrogative.* Burlington, VT: Ashgate, 2002. xi, 272 p. Cloth $79.95.

SOULE, EDWARD: *Morality and Markets: The Ethics of Government Regulations.* Lanham, MD: Rowman and Littlefield, 2003. xiv, 193 p. Cloth $70.00, paper $23.95.

SPADER, PETER H.: *Scheler's Ethical Personalism: Its Logic, Development, and Promise.* New York: Fordham, 2002. xviii, 327 p. Cloth $45.00, paper $20.00.

SPELMAN, ELIZABETH V.: *Repair: The Impulse to Restore in a Fragile World.* Boston: Beacon, 2002. 165 p. Cloth $24.00.

STAMOS, DAVID N.: *The Species Problem: Biological Species, Ontology, and the Metaphysics of Biology.* Lanham, MD: Lexington, 2003. 381 p. Cloth $80.00.

STANISTREET, PAUL: *Hume's Scepticism and the Science of Human Nature.* Burlington, VT: Ashgate, 2002. xi, 226 p. Cloth $69.95.

STECKER, ROBERT: *Interpretation and Construction: Art, Speech, and the Law.* Malden, MA: Blackwell, 2003. x, 212 p. Paper $27.95.

STERELNY, KIM: *Thought in a Hostile World: The Evolution of Human Cognition.* Malden, MA: Blackwell, 2003. xi, 262 p. Cloth $35.00.

STOKHOF, MARTIN: *World and Life as One: Ethics and Ontology in Wittgenstein's Early Thought.* Stanford: University Press, 2002. xvi, 329 p. Cloth $60.00.

STONEHAM, TOM: *Berkeley's World: An Examination of the Three Dialogues.* New York: Oxford, 2002. xviii, 308 p. Cloth $65.00, paper $21.95.

STREVENS, MICHAEL: *Bigger than Chaos: Understanding Complexity through Probability.* Cambridge: Harvard, 2003. xii, 413 p. Cloth $59.95.

SULLIVAN, STEFAN: *Marx for a Post-Communist Era: On Poverty, Corruption, and Banality.* New York: Routledge, 2002. x, 191 p. Cloth $80.00, paper $24.95.

SWANTON, CHRISTINE: *Virtue Ethics: A Pluralistic View.* New York: Oxford, 2003. xi, 312 p. Cloth $35.00.

SWINBURNE, RICHARD: *The Resurrection of God Incarnate.* New York: Oxford, 2003. vii, 224 p. Cloth $70.00, paper $24.95.

TANTILLO, ASTRIDA ORLE: *The Will to Create: Goethe's Philosophy of Nature.* Pittsburgh: University Press, 2002. xii, 241 p. Cloth $32.50.

TAYLOR, CRAIG: *Sympathy: A Philosophical Analysis.* New York: Palgrave Macmillan, 2002. ix, 155 p. Cloth $55.00.

TAYLOR, RICHARD: *Virtue Ethics: An Introduction.* Amherst, NY: Prometheus, 2002. xvi, 125 p. Paper $18.00.

TEMES, PETER S.: *The Just War: An American Reflection on the Morality of War in Our Time.* Chicago: Ivan R. Dee, 2003. xv, 219 p. Cloth $25.00.

THORNHILL, CHRIS: *Karl Jaspers: Politics and Metaphysics.* New York: Routledge, 2002. x, 272 p. Cloth $95.00.

TIMMONS, MARK: *Moral Theory: An Introduction.* Lanham, MD: Rowman and Littlefield, 2002. x, 291 p. Cloth $65.00, paper $18.95.

TOULMIN, STEPHEN E.: *The Uses of Argument,* Updated Version. New York: Cambridge, 2003. xiv, 247 p. Cloth $55.00, paper $20.00.

TUCKNESS, ALEX: *Locke and the Legislative Point of View: Toleration, Contested Principles, and the Law.* Princeton: University Press, 2002. xiii, 206 p. Cloth $49.50, paper $17.95.

TURNER, STEPHEN P.: *Brains/Practices/Relativism.* Chicago: University Press, 2002. ix, 214 p. Paper $19.00.

UMPHREY, STEWART: *Complexity and Analysis.* Lanham, MD: Lexington, 2002. vii, 351 p. Cloth $90.00, paper $26.95.

UNGER, PETER: *Philosophical Relativity.* New York: Oxford, 2003. x, 132 p. Paper $19.95.

URBINATI, NADIA: *Mill on Democracy from the Athenian Polis to Representative Government.* Chicago: University Press, 2002. xii, 293 p. Cloth $37.50.

VALLICELLA, WILLIAM F.: *A Paradigm Theory of Existence: Onto-theology Vindicated.* Boston: Kluwer, 2002. xii, 281 p. Cloth $91.00.

VAN FRAASSEN, BAS C.: *The Empirical Stance.* New Haven, CT: Yale, 2002. xix, 282 p. Cloth $30.00.

VAN INWAGEN, PETER: *Metaphysics,* Second Edition. Boulder, CO: Westview, 2002. xi, 235 p. Cloth $82.00, paper $26.00.

VAUGHN, LEWIS, and AUSTIN DACEY: *The Case for Humanism: An Introduction.* Lanham, MD: Rowman and Littlefield, 2003. ix, 223 p. Cloth $75.00, paper $26.95.

VELKLEY, RICHARD L: *Being after Rousseau: Philosophy and Culture in Question.* Chicago: University Press, 2002. x, 192 p. Cloth $40.00, paper $18.00.

VERBEEK, BRUNO: *Instrumental Rationality and Moral Philosophy: An Essay on the Virtues of Coorperation.* Boston: Kluwer, 2002. xii, 286 p. Cloth $82.00.

VERENE, DONALD PHILLIP: *The Art of Humane Education.* Ithaca, NY: Cornell, 2002. xiii, 80 p. Cloth $19.95; *Knowledge of Things Human and Divine: Vico's New Science and Finnegans Wake.* New Haven, CT: Yale, 2003. xiv, 264 p. Cloth $45.00.

VOGLER, CANDACE: *Reasonably Vicious.* Cambridge: Harvard, 2002. viii, 295 p. Cloth $45.00.

WALDRON, JEREMY: *God, Locke, and Equality: Christian Foundations of John Locke's Political Thought.* New York: Cambridge, 2002. xii, 263 p. Cloth $60.00, paper $22.00.

WALTON, DOUGLAS: *Legal Argumentation and Evidence.* University Park: Penn State UP, 2002. xvii, 374 p. Cloth $65.00; *Relevance in Argumentation.* Mahwah, NJ: Erlbaum, 2004. xvi, 311 p. Cloth $69.95, paper $34.50.

WARNOCK, MARY: *Making Babies: Is There a Right Way to Have Children*? New York: Oxford, 2002. vi, 120 p. Cloth $13.95, paper $12.95.

WARREN, JAMES: *Epicurus and Democritean Ethics: An Archaeology of Ataraxia.* New York: Cambridge, 2002. xiv, 241 p. Cloth $55.00.

WEATHERSTON, MARTIN: *Heidegger's Interpretation of Kant: Categories, Imagination and Temporality.* New York: Palgrave Macmillan, 2002. xi, 209 p. Cloth $62.00.

WEGNER, DANIEL M.: *The Illusion of Conscious Will.* Cambridge: MIT, 2002. xi, 405 p. Cloth $34.95.

WEHANE, PATRICIA H., and TARA J. RADIN, with NORMAN E. BOWIE: *Employment and Employee Rights.* Malden, MA: Blackwell, 2003. x, 228 p. Paper $29.95.

WEISS, BERNHARD: *Michael Dummett.* Princeton: University Press, 2002. ix, 197 p. Cloth $55.00, paper $18.95.

WEISS, PAUL: *Surrogates.* Bloomington: Indiana UP, 2002. xviii, 181 p. Cloth $34.95.

WEISSMAN, DAVID: *Lost Souls: The Philosophic Origins of a Cultural Dilemma.* Albany: SUNY, 2003. xiii, 210 p. Cloth $57.50, paper $18.95.

WEITHMAN, PAUL J.: *Religion and the Obligations of Citizenship.* New York: Cambridge, 2002. xi, 227 p. Cloth $55.00.

WESTPHAL, KENNETH R.: *Hegel's Epistemology: A Philosophical Introduction to the Phenomenology of Spirit.* Indianapolis: Hackett, 2003. xvi, 146 p. Cloth $34.95, paper $12.95.

WILLIAMS, BERNARD: *Truth and Truthfulness: An Essay in Genealogy.* Princeton: University Press, 2002. xi, 328 p. Cloth $27.95.

WINCHESTER, JAMES J.: *Aesthetics across the Color Line: Why Nietzsche (Sometimes) Can't Sing the Blues.* Lanham, MD: Rowman and Littlefield, 2002. xi, 145 p. Cloth $60.00, paper $19.95.

WINGO, AJUME H.: *Veil Politics in Liberal Democratic States.* New York: Cambridge, 2003. xvii, 162 p. Cloth $55.00, paper $22.00.

WIRTH, JASON M.: *The Conspiracy of Life: Meditations on Schelling and His Time.* Albany: SUNY, 2003. ix, 290 p. Cloth $71.50, paper $23.95.

WISNOVSKY, ROBERT: *Avicenna's Metaphysics in Context.* Ithaca, NY: Cornell, 2003. viii, 305 p. Cloth $65.00.

WITT, CHARLOTTE: *Ways of Being: Potentiality and Actuality on Aristotle's Metaphysics.* Ithaca, NY: Cornell, 2003. x, 161 p. Cloth $35.00.

WOLFF, JONATHAN: *Why Read Marx Today?* New York: Oxford, 2002. viii, 136 p. Cloth $20.00.

WOODS, JOHN: *Paradox and Paraconsistency: Conflict Resolution in the Abstract Sciences.* New York: Cambridge, 2003. xviii, 362 p. Cloth $70.00, paper $26.00.

WRIGHT, CRISPIN: *Saving the Differences: Essays on Themes from Truth and Objectivity.* Cambridge: Harvard, 2003. ix, 549 p. Cloth $55.00.

YOUNG-AH GOTTLIEB, SUSANNAH: *Regions of Sorrow: Anxiety and Messianism in Hannah Arendt and W.H. Auden.* Stanford: University Press, 2003. xii, 306 p. Cloth $65.00, paper $24.95.

YURACKO, KIMBERLY A.: *Perfectionism and Contemporary Feminist Values.* Bloomington: Indiana UP. x, 170 p. Cloth $44.95, paper $19.95.

YUSA, MICHIKO: *Zen and Philosophy: An Intellectual Biography of Nishida Kitaro.* Honolulu: Hawaii UP, 2002. xxvi, 482 p. Paper $29.95.

ZAHAVI, DAN: *Husserl's Phenomenology.* Stanford: University Press, 2003. viii, 182 p. Cloth $45.00, paper $15.95.

ZIPORYN, BROOK: *The Penumbra Unbound: The Neo-Taoist Philosophy of Guo Xiang.* Albany: SUNY, 2003. ix, 186 p. Cloth $68.50, paper $22.95.

ŽIŽEK, SLAVOJ: *For They Know Not What They Do: Enjoyment as a Political Factor,* Second Edition. New York: Verso, 2002. cvii, 288 p. Paper $21.00; *The Puppet and the Dwarf: The Perverse Core of Christianity.* Cambridge: MIT, 2003. viii, 188 p. Paper $16.95.

ZUCKERT, MICHAEL P.: *Launching Liberalism: On Lockean Political Philosophy.* Lawrence: Kansas UP, 2002. xi, 375 p. Cloth $40.00, paper $19.95.

ZUPKO, JACK: *John Buridan: Portrait of a Fourteenth-century Arts Master.* Notre Dame: University Press, 2003. xix, 446 p. Cloth $70.00, paper $40.00.

NEW BOOKS: PAPERBACK REPRINTS

ARON, RAYMOND: *The Dawn of Universal History: Selected Essays from a Witness of the Twentieth Century.* YAIR REINER, ed., BARBARA BRAY, trans. New York: Basic Books, 2002/2003. xxvi, 518 p. Paper $22.00.

BECKER, CARL L.: *The Heavenly City of the Eighteenth-Century Philosophers, Second Edition.* New Haven: Yale, 1932/2003. xix, 168 p. Paper $16.00.

BILLINGTON, RAY: *Living Philosophy: An Introduction to Moral Thought,* Third Edition. New York: Routledge, 1988/2003. xviii, 354 p. Paper $27.95.

BROUGHTON, JANET: *Descartes's Method of Doubt.* Princeton: University Press, 2002/2003. xv, 217 p. Paper $16.95.

BROWN, JAMES ROBERT: *Who Rules in Science?* Cambridge: Harvard, 2001/2004. xiii, 236 p. Paper $16.95.

CHATTERJEA, TARA: *Knowledge and Freedom in Indian Philosophy.* Lanham, MD: Lexington, 2002/2003. xvi, 161 p. Paper $24.95.

DAVIDSON, ARNOLD I.: *The Emergence of Sexuality: Historical Epistemology and the Formation of Concepts.* Cambridge: Harvard, 2001/2004. xvi, 254 p. Paper $19.95.

GEUSS, RAYMOND: *Public Goods, Private Goods.* Princeton: University Press, 2001/2003. xxvi, 148 p. Paper $14.95.

HADOT, PIERRE: *What Is Ancient Philosophy?* MICHAEL CHASE, trans. Cambridge: Harvard, 2002/2004. xii, 362 p. Paper $15.95.

HAKSAR, VINIT: *Rights, Communities, and Disobedience: Liberalism and Gandhi.* New York: Oxford, 2001/2003. xxxi, 202 p. Paper $14.95.

KANT, IMMANUEL: *Observations on the Feeling of the Beautiful and Sublime.* JOHN T. GOLDTHWAIT, trans. Berkeley: California UP, 1991/2003. 124 p. Paper $16.95.

LACOCQUE, ANDRÉ, and PAUL RICOEUR: *Thinking Biblically: Exegetical and Hermeneutical Studies.* DAVID PELLAUER, trans. Chicago: University Press, 1998/2003. xix, 441 p. Paper $18.00.

LAMPERT, LAURENCE: *Nietzsche's Task: An Interpretation of Beyond Good and Evil.* New Haven: Yale, 2001/2004. x, 320 p. Paper $24.00.

LOVIBOND, SABINA: *Ethical Formation.* Cambridge: Harvard, 2002/2004. xv, 203 p. Paper $19.95.

MARGALIT, AVISHAI: *The Ethics of Memory.* Cambridge: Harvard, 2002/2004. xi, 227 p. Paper $14.95.

O'HAGAN, TIMOTHY: *Rousseau.* New York: Routledge, 1999/2003. xvi, 320 p. Paper, $27.95.

PUTNAM, HILARY: *The Collapse of the Fact/Value Dichotomy and Other Essays.* Cambridge: Harvard, 2002/2004. x, 190 p. Paper $16.95.

RICHARDSON, HENRY S.: *Democratic Autonomy: Public Reasoning about the Ends of Policy.* New York: Oxford, 2002/2004. xii, 316 p. Paper $19.95.

RICOEUR, PAUL: *The Just.* DAVID PELLAUER, trans. Chicago: University Press, 2000/2003. xxiv, 161 p. Paper $14.00.

SCHEFFLER, SAMUEL: *Boundaries and Allegiances: Problems of Justice and Responsibility in Liberal Thought.* New York: Oxford, 2001/2002. 221 p. Paper $19.95.

SCRUTON, ROGER: *A Short History of Modern Philosophy: From Descartes to*

Wittgenstein, Second Edition. New York: Routledge, 1995/2002. ix, 318 p. Paper $14.95.

SELIGMAN, ADAM B.: *Modernity's Wager: Authority, the Self, and Transcendence.* Princeton: University Press, 2000/2003. xii, 177 p. Paper $19.95.

SEN, AMARTYA: *Rationality and Freedom.* Cambridge: Harvard, 2004. ix, 736 p. Paper $19.95.

SHAND, JOHN: *Philosophy and Philosophers: An Introduction to Western Philosophy.* Montreal: McGill-Queen's, 1993/2002. x, 341 p. Paper $19.95.

SMART, BARRY: *Michel Foucault, Revised Edition.* New York: Routledge, 1985/2002. xv, 150 p. Paper $14.95.

STEINER, MARK: *The Applicability of Mathematics as a Philosophical Problem.* Cambridge: Harvard, 1998/2002. viii, 215 p. Paper $19.95.

TAMEN, MIGUEL: *Friends of Interpretable Objects.* Cambridge: Harvard, 2001/2004. 194 p. Paper $16.95.

TAYLOR, CHARLES: *Varieties of Religion Today: William James Revisited.* Cambridge: Harvard, 2002/2003. vii, 127 p. Paper $12.00.

TAYLOR, MARK C.: *The Moment of Complexity: Emerging Network Culture.* Chicago: University Press, 2001/2003. xii, 340 p. Paper $17.00.

VAN DEN BERG, AXEL: *The Immanent Utopia: From Marxism on the State to the State of Marxism.* New Brunswick, NJ: Transaction, 1989/2003. xli, 580 p. Paper $29.95.

WELTON, DONN: *The Other Husserl: The Horizons of Transcendental Phenomenology.* Bloomington: Indiana UP, 2000/2002. xvi, 496 p. Paper $29.95.

WIERENGA, EDWARD R.: *The Nature of God: An Inquiry into Divine Attributes.* Ithaca, NY: Cornell, 1989/2003. xii, 238 p. Paper $24.95.

WILLIAMS, BERNARD: *Truth and Truthfulness: An Essay in Genealogy.* Princeton: University Press, 2002/2004. xi, 328 p. Paper $17.95.

WOODRUFF, PAUL: *Reverence: Renewing a Forgotten Virtue.* New York: Oxford, 2001/2002. x, 248 p. Paper $11.95.

HEAVEN'S FRACTAL NET

Retrieving Lost Visions in the Humanities

William J. Jackson

Explores the beautiful and intriguing fractal patterns found in art, literature, religion, and beyond.

cloth $59.95 paper $29.95

JOHN DEWEY AND MORAL IMAGINATION

Pragmatism in Ethics

Steven Fesmire

A broad pragmatist grounding that revitalizes ethics and moral philosophy.

cloth $49.95 paper $19.95

INDIANA UNIVERSITY PRESS

at bookstores • Order: 1-800-842-6796 • iupress.indiana.edu

Studies in Continental Thought

John Sallis, Editor

THINKING THROUGH FRENCH PHILOSOPHY

The Being of the Question

Leonard Lawlor

A powerful and original engagement with France's most influential philosophers.

cloth $59.95 paper $24.95

RACE AND RACISM IN CONTINENTAL PHILOSOPHY

Edited by Robert Bernasconi with Sybol Cook

New thinking on how Continental philosophers approach race and racism.

cloth $59.95 paper $24.95

TRUTH AND GENESIS

Philosophy as Differential Ontology

Miguel de Beistegui

Develops a new philosophy that is grounded in both artistic and scientific truth.

cloth $60.00 paper $29.95

THINKING WITH HEIDEGGER

Displacements

Miguel de Beistegui

An invitation to think with Heidegger and draw his ideas out into a more contemporary idiom.

cloth $59.95 paper $24.95

HEIDEGGER'S *CONTRIBUTIONS TO PHILOSOPHY*

An Introduction

Daniela Vallega-Neu

Short, clear, readable introduction to Heidegger's second most important work.

cloth $39.95 paper $19.95

RETHINKING NATURE

Essays in Environmental Philosophy

Edited by Bruce V. Foltz and Robert Frodeman

A landmark volume for defining the terrain of environmental philosophy.

cloth $59.95 paper $24.95

SCHELLING NOW

Contemporary Readings

Edited by Jason M. Wirth

Places Schelling at the core of contemporary philosophical debates.

cloth $49.95 paper $22.95

THE NEW KIERKEGAARD

Edited by Elsebet Jegstrup

New essays on Kierkegaard that read his philosophy in the light of deconstruction.

cloth $59.95 paper $24.95

THE NEW HUSSERL

A Critical Reader

Edited by Donn Welton

Presents fresh research and new perspectives on Husserl's thought.

cloth $75.00 paper $29.95

THE JOURNAL OF PHILOSOPHY

Celebrates 100 Years

Special Limited
Time Offer
Full Volumes 1-99
1904-2002

$50.00 per Volume

All volumes are available for $50.00
plus $5.00 shipping and handling.
Send check or money order to:

The Journal of Philosophy
1150 Amsterdam Avenue, MC 4972
Columbia University
New York, NY 10027
(212) 666-4419
www.journalofphilosophy.org