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

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

Florida Philosophical Review: The Journal of the Florida Philosophical Association (ISSN 1535-3656) is published twice a year (in June and December) by the University of Central Florida Department of Philosophy, which assumes no responsibility for statements expressed by the authors. Copyright on *Florida Philosophical Review* is held by the University of Central Florida. However, authors hold copyright privileges on individual essays published under their names. As an electronic journal, *Florida Philosophical Review* is distributed free of charge on-line. Individuals desiring a CD version of a volume of the journal may purchase it for $18.00 ($10.00 for single issues). The institutional price for one year (two issues) is $50.00.

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

EDITORS

Shelley M. Park
Associate Professor and Chair,
Department of Philosophy
University of Central Florida
Orlando, FL 32816-1352

Nancy A. Stanlick
Assistant Professor,
Department of Philosophy
University of Central Florida
Orlando, FL 32816-1352

EDITORS' INTRODUCTION

This issue of the *Florida Philosophical Review* marks the beginning of our second year of publication. In it are selected papers and symposia from the 47th annual meeting of the Florida Philosophical Association held at Stetson University in DeLand, Florida on November 9th and 10th, 2001.

Kirk Ludwig's Presidential Address to the members of the FPA, "The Arrangement of the Soul: Philosophy and the Professional Philosopher," opens this volume with thoughts on the work and role of the professional philosopher. Ludwig addresses the paradoxical nature of our engagement in the study and pursuit of the philosophical life such that, as professional philosophers, we find ourselves forced into specialization and compartmentalization, leading us away from the pursuit and attainment of synoptic understanding that first drew us to philosophy. Given our myriad institutional duties, the specialization of our philosophical sub-fields, and the resulting compartmentalization of our professional from our private lives, it is increasingly difficult for the professional philosopher to live an examined life. Thus, the central question of Ludwig's address is a question central to all of us (although we may too infrequently reflect on it), namely: "How can we retain our allegiance to philosophy while being professional philosophers?"

Following the Presidential Address are this year's prize-winning student papers. Jeremy Kirby's paper, "Aristotle and Supervenience Physicalism," won the 2001 Outstanding Graduate Philosophy Paper award. In this essay, Kirby argues—against scholars such as Michael Wedin and Victor Caston—that Aristotle's work is not friendly to supervenience physicalism. He rests his argument not only on evidence found in Aristotle's *Physics*, but also on evidence from Aristotle's psychological and biological works. A primary impetus for attributing supervenience physicalism to Aristotle concerns certain apparent contradictions in the Aristotelean texts (e.g. Aristotle says that φ-states are generated and that φ-states are not generated) allegedly resolved by interpretations friendly to supervenience physicalism. Kirby concludes that "there are means more benign . . . for dealing with [such] apparent contradictions" than those pursued in defense of the thesis that Aristotle was friendly to supervenience physicalism.

David Barnett of New College of Florida is the winner of the FPA's 2001 Gerrit and Edith Schipper Undergraduate Award for his paper, "Hempel on Intertheoretic Reduction." Barnett argues that Hempel's contention that all biological phenomena can be explained by physical processes is flawed and therefore does not lend itself to the unity of science. "Using the morals . . . [drawn] from [Hempel's] failures," Barnett concludes by suggesting "rough outlines of some alternative accounts of intertheoretic reduction."

Two symposia consisting of discussions of new books by Florida philosophers close this issue of the *Florida Philosophical Review*. Symposium I presents critical commentary on Martin Schönfeld's *The Philosophy of the Young Kant: The Precritical Project* (Oxford: Oxford UP, 2000). In the opinions of the commentators, this is a ground-breaking and significant work shedding light on Kant's early philosophical development and illuminating aspects of his mature philosophy. Jennifer Uleman (University of Miami), Byron Williston (Wilfrid Laurier University) and Sidney Axinn (University of South Florida; Temple University, Emeritus) offer insights into Schönfeld's work.

Jennifer Uleman summarizes succinctly the valuable elements of Schönfeld's book both for historians of philosophy and for those who simply wish to understand more fully the work of Kant himself and poses four questions regarding Kant's precritical work. Her questions concern the young Kant's (seemingly) teleological views of the will and his resolution to the problem of determinism and freedom, as well as more general issues regarding the import of Kant's precritical works for understanding and assessing his critical works, and the relationship of the history of philosophy to philosophy itself. Byron Williston addresses two particular interpretations that Schönfeld makes of the young Kant, arguing, first, that Kant may not need to be interpreted as engaging in *self*-critique in the *Dreams of a Spirit Seer* and, secondly, that Kant's precritical works may already contain the view that autonomy is to be grounded in practical judgement. Lastly, Sidney Axinn takes a look at Schönfeld's work from the point of view of its value to the history of ideas and poses questions regarding Kant's early views on negation and possibility.

In responding to his commentators' remarks and queries, Schönfeld highlights a major thesis of *The Philosophy of the Young Kant,* namely, that Kant's precritical works were and continue to be valuable contributions to the history of ideas. According to Schönfeld, "the early Kant was an original and innovative thinker wrestling with timely issues and perennial questions, who systematically constructed an ambitious reconciliation of science and metaphysics." Although the precritical project was ultimately a failure for Kant, Schönfeld contends, the young Kant's *oeuvre* sheds light on the motivations and rationales of the mature Kant's critical philosophy—in addition to having merits of its own.

In Symposium II, Risto Hilpinen (University of Miami), Crystal Thorpe (University of Miami), and Peter Dalton (Florida State University) comment on Alfred Mele's book, *Self-Deception Unmasked* (Princeton and Oxford: Princeton UP, 2001). In this work, Mele offers a unique analysis of the philosophical, social, and psychological aspects of the phenomenon of self-deception in which a person in possession of adequate or sufficient evidence for the truth of proposition *p* is able to convince himself of the truth of *not p*. Mele offers an analysis of self-deception that promises to be a springboard for further fruitful and fascinating discussion of which this symposium is an interesting and thought-provoking contribution.

Peter Dalton argues that a conceptual analysis of self-deception would need to include as a necessary condition that a self-deceived person is unaware that he reasoned incorrectly in arriving at his false belief. Although this adds a condition to Mele's own stated criteria for self-deception, Dalton argues, it also provides us with "a strong reason for agreeing with Mele that explanations of self-deception must stress non-agency." Like Dalton, Crystal Thorpe agrees with Mele that self-deception is not adequately captured by the "traditional view," namely, the view that depicts self-deception as intentional and as involving the holding of contradictory beliefs. However, Thorpe suggests that a rejection of such a view is warranted for reasons that go beyond Mele's own: it is not merely that empirical psychology shows such a view wrong-headed; it is, according to Thorpe, that "the traditional view fails to capture our attitudes toward self-deceived people." In the final commentary, Risto Hilpinen situates Mele's discussion of self-deception within the context of historical literature on the topic, including work by Jean Paul Sartre, G.E. Moore, and Imre Lakatos. Of particular interest to Hilpinen are the links between self-deception and motivational bias and between self-deception and confirmation bias, as discussed by Mele. Hilpinen's use of Imre Lakatos's conception of scientific methodology raises an interesting analogy between degenerative research programs and garden-variety cases of self-deception.

In his responses to these commentaries, Mele discusses empirical evidence for the confirmation bias and extreme cases of self-deception, arguing that even an extreme case of self-deception can be explained by empirically intelligible phenomena simpler than the phenomena countenanced by, say, a Freudian model of self-deception. Thus, such cases do "not require the machinery of a traditional conception of self-deception—that is, simultaneously believing that *p* and believing that ~*p* and intentionally bringing it about that one acquires the belief that one favors." He further attempts to imagine the form that a Dalton-type counter-example to his analysis of self-deception might take, concluding that formulating such a counter-example is likely impossible.

Collectively, the essays and symposia included in this volume embody an array of philosophical interests and methods. The work included here addresses texts and issues in ancient, modern, and contemporary philosophy, reflects both philosophical and metaphilosophical concerns, and demonstrates a commitment to techniques of translation and interpretation, logical analysis and argumentation, and attention to relevant empirical evidence for (and against) philosophical positions.

Florida Philosophical Review invites the submission of papers and book reviews from philosophers with varied philosophical interests for review and consideration for inclusion in upcoming issues. Our next issue (Volume II, Issue 2) will be devoted primarily to papers and commentaries on the topic of terrorism. Among topics for consideration are: the problem of evil; the distinction between individual and corporate (state, organizational) responsibility; desert and the limits of punishment; conceptions of justice or just war; conceptual analyses of terrorism; possible tensions between liberty and security; responsibilities of the media; cultural analyses of media coverage and political rhetoric; peace and reconciliation; understanding the 'Other;' conceptions of rationality; theoretical and practical issues related to patriotism and nationalism; feminist and post-colonialist analyses of conflict, responsibility and otherness. We also welcome papers on other issues of relevance from epistemological, ethico-political, socio-cultural, and general philosophical perspectives. The deadline for submissions for this issue is August 1, 2002. Please see the call for papers and the invitation for book reviews included in this issue and appearing on the main *FPR* website.

Volume III, Issue 1 (June 2003) will include selected papers from the 2002 Florida Philosophical Association conference and Volume III, Issue 2 (December 2003) will include graduate student articles on any area or aspect of philosophical inquiry. Submissions of papers and book reviews for these and all other issues of the journal are welcome from philosophers both from within and without the state of Florida. Our general, meeting, and special issues are all open to papers on a specific issue's theme as well as other topics of philosophical interest and relevance.

We hope that you will continue to be a regular reader of and contributor to *Florida Philosophical Review: The Journal of The Florida Philosophical Association.*

Shelley Park and Nancy Stanlick, Editors
Florida Philosophical Review: The Journal of The Florida Philosophical Association.
June 30, 2002

The Arrangement of the Soul: Philosophy and the Professional Philosopher

Presidential Address of the 47th Annual Meeting
of the Florida Philosophical Association

***Kirk Ludwig**, University of Florida*

I have approached the task of giving this Presidential Address with some trepidation. It is, in any circumstance, a matter of some difficulty, occurring on the evening of a day full of long talks, when the last thing anyone wants to hear is more of the same. It occurs additionally not before but after the banquet, after the food, after the pie and cake, after . . . the wine. But worst of all, this presidential address occurs after the bravura performance of last year's president, Aron Edidin, who delivered his address in iambic pentameter. His parting words, delivered in a rhymed couplet, were:

Now nears my end of presidential work,
I sigh relief, and pass the torch to Kirk.

I fear I will be burned by the torch that has been passed to me. My colleague Bob D'Amico told me that the only way I could top that would be by delivering my address in Terza Rima. I confess myself, however, wholly unable to rise to the challenge. At least I can console myself with the thought that I will make the task of our next year's president, Martin Schönfeld, an easier one.

What I will do by way of compensation for not being able to deliver my remarks in Terza Rima is to make them relatively brief. I had thought, momentarily, of giving a talk on semantic vagueness. For I have a curious proof, with my colleague Greg Ray, that nothing that I say, or that you say, is false. That's the good news. The bad news is that it is not true either, insofar as what you say involves the use of vague terms. But on reflection this seemed more likely to be a recipe for inducing vagueness, given the setting, than for clarifying it. I am therefore going to talk about something that is not a technical issue in philosophy, and about something on which I am certainly in no sense an expert. I wish I were.

It is a question on which I have been reflecting recently, for a variety of reasons, and one which I think we all think about when we have time, and, in part, because we seldom have time. I will tell you at the outset that I will not give you an answer to the question. It is, in part, a kind of practical question. We need perhaps to find a *phronimos*, the man or woman of practical wisdom, and I am not that. If raising the question without answering it has any virtue, it will be from its prompting additional reflection and discussion.

We are professional philosophers—academic philosophers. We hold teaching jobs, we write papers and books, for professional journals and academic presses, for other professional philosophers. We edit collections, read the papers and books of our brethren. We perform administrative tasks for our departments and institutions. It is a job. It is a profession, with all the trappings of a profession. It is also a strange way to be a philosopher. And it is that paradox (to speak loosely), the paradox of being a professional philosopher, about which I want to talk.

Why do I speak of it as a paradox? I have two things in mind. I will approach them indirectly.

When I was a graduate student, I overheard one of my teachers, Barry Stroud, wishing one day that he felt more like a philosopher and less like an employee of the state. I am not quite sure what he was thinking, but I think I have more sympathy for his remark now than I did then. (I think at the time I must have been trying to make an appointment with him.)

I am not sure I want to say in general why we become interested in philosophy or what draws us into it. I will confine myself to some autobiographical remarks. As an undergraduate, I was a physics major. I made a choice at the outset of my college career to study physics as opposed to English because I wanted to know as much as I could about . . . as much as I could, and I thought if I did not study the sciences early, it would not be practical to return to it in a serious manner later on. And physics I conceived of as the most fundamental of the sciences. It was. I do not regret my decision. It was clear to me, though, that pursuing graduate school in physics was not what I wanted to do. If you pursue graduate work in physics, you specialize: you learn more and more about . . . less and less. I wanted to learn more and more about more and more. Some minimal exposure to philosophy suggested to me that this was a subject that would afford me the kind of freedom to think about whatever interested me that I wanted, and at the level of generality and abstraction at which I wanted to think about it. Though I was at the time terribly ignorant about philosophy, about this I was right. I worked for a couple of years, took some graduate courses in both physics and philosophy, and went to Berkeley to study philosophy, not physics.

I spent seven years at Berkeley. I was poor, but I was rich in time and freedom. Now the balance has shifted in the other direction. This is, in part, what I think Barry Stroud had in mind in expressing a desire to feel less like an employee of the state. Our jobs, our professional duties as teachers and administrators, and as philosophers, do not in fact provide us with the kind of time and freedom which is necessary for pursuing philosophy as we would like, and for pursuing philosophy as it should be pursued. But this is only part of it and, in a sense, the less important part. I enjoy each of the things I do individually, even administrative work: it is just that it is an embarrassment of riches.

There is another aspect of the professionalization of philosophy, however, which limits our freedom in a different way.

The professionalization of philosophy, and of one's philosophical work, in consequence, means that one is responsible to a professional literature, a specialist literature. One's work is (typically) confined to official organs of the profession. One is expected by referees to pay attention to what people have been paying attention to. One's institution expects one to contribute regularly to the official organs of one's profession and, indeed, as a new assistant professor one's professional life depends upon conforming to these unwritten canons.

Now, what is wrong with all of this? Of course, I do not foolishly protest against the inevitable specialization that attends the advancement of any field. It is unavoidable and necessary for progress.

Yet there are some dangers here as well. And we fail to note them at our peril as philosophers.

There are familiar vices attendant on these institutional arrangements. There is quite a bit more published than deserves to be published. In general, the profession would benefit from its members publishing less of higher quality. There is so much noise in the journals that it is difficult to find what is worth listening to, and too often people listen to the same voices over and over again simply because they trust them and don't have the time to sort the wheat from the chaff. This accounts, in part, for what sometimes strikes me as the boringly limited range of things people find to write about in philosophy. (Everyone writes about what a few write about. Enough, for example, on proper names already!) Perhaps it also accounts for certain persistent confusions, which have been cleared up, but about which not everyone, not even all journal referees, have got the word (you no doubt can find your own examples in your field of expertise). Younger members of the profession in particular often find themselves forced to place their work in print before it is fully developed in order to provide external evidence of professional stature. There is a limited amount we can do about these things. They are necessary accompaniments of the procedures we put in place to evaluate members of the profession and their work—itself a necessary if unpleasant feature of having a profession at all. To borrow a phrase from Samuel Johnson, which he made in reference to notes: they are necessary, but they are necessary evils.

But these features attend every academic profession. The thing I have in mind is a kind of internal tension between the necessity of this kind of professionalization in philosophy and what it is we seek for ourselves in philosophy—what it is to be philosophers. And it strikes me that there is not, or is not the necessity of, the same kind of tension in other academic disciplines.

For example, science is a collaborative enterprise and must be by its very nature. The specialist contributes to a body of knowledge that is accumulated by the community, and that is her proper role. The enterprise of science tries to reach for synoptic understanding, but its under workers need not be doing that in doing what they do. Philosophy seeks a more synoptic understanding than science does. But, it seems to me, in contrast, it is not just the goal of the enterprise of philosophy to seek a synoptic understanding of ourselves and our world and our relation to it and to each other, but of each one of us as philosophers. It is part of the impulse to engage in philosophical reflection to seek that kind of synoptic understanding. It was certainly the impulse I followed. When we become professional philosophers, when we become, in a word, specialists, we bind ourselves to a discipline that prevents us from (or threatens to frustrate us in) pursuing what it was that got us into philosophy in the first place (or got me into philosophy in the first place). So this is the first tension, the first aspect of what I called a paradox, in being a professional philosopher. Our professional energies are directed into a relatively small range of problems or areas. But our—if I can put it like this—professional aspirations cannot but be frustrated by this. (There are people who defy this pressure, of course, but it is a pressure nonetheless.)

I turn now to a different product of professionalization in philosophy, which also seems to me to pull against how we think about the role of philosophy in our lives as philosophers. It is connected with the first point and I will come back to that in a moment.

Another teacher of mine, when I was still newer to philosophy, told me that professional philosophers by and large compartmentalized their professional and private lives. It struck me as strange at the time, for there was then no line in my life on the other side of which there was something besides the things I was interested in understanding. Even now, I must say, there is not a lot on the other side of that line. But I have a vastly improved understanding of the force of that remark nonetheless.

Professionalization encourages a certain kind of compartmentalization of one's philosophical life from one's life in general. I don't say this happens for everyone, but I think the pressure is there for everyone who comes to philosophy with something like the ideal of philosophy that informed the philosophers of ancient Greece. If we think of becoming a philosopher as adopting a way of life, not just a profession, then to become a professional philosopher is to that extent to turn away from being a philosopher, and our becoming professional philosophers turns us away from thinking about philosophy as a way of life. A profession is not a way of life—not anymore, at any rate—it is a way of earning a living.

What do I mean in talking about becoming a philosopher being a matter of adopting a way of life? This is connected with what I think of as philosophy's aim for synoptic understanding. Unlike other academic fields, philosophy is not characterized by a subject matter that limits its concerns, but by a concern for foundations in every field of inquiry and every category of human activity, and by a concern for their interconnections. The synoptic understanding we seek includes an understanding of ourselves and our lives, our relations to others, and to the social, economic and political institutions within which we live our lives. If we think of the philosopher as someone who seeks this synoptic understanding, then part of what is involved in becoming a philosopher involves a concern to extend one's examination of things to one's life as a whole. One's life, and the way one lives it, then, must be informed by this examination. I do not know whether the unexamined life is not worth living, but we might say that it is not the life of a philosopher. One of the things we surely seek, in the journey from dark to dark, in the arc of a life, is what its proper shape should be, and so philosophy if pursued thoroughly must have a practical dimension.

In Book X of the *Republic* (which tells us to our delight that we are most like the divine when we pursue philosophy—no wonder we assign it to our students), Plato has Socrates telling the story of "a brave Pamphylian man called Er, the son of Armenias" (614b), who journeys to the world of the dead and returns. He tells the story of the souls in Hades being recycled from death to life, given a choice by the Fates of what lives they will lead:

> Here is the message of Lachesis, the maiden daughter of necessity: 'Ephemeral souls, this is the beginning of another cycle that will end in death. Your daimon or guardian spirit will not be assigned to you by lot; you will choose him. The one who has the first lot will be the first to choose a life to which he will then be bound by necessity. Virtue knows no master; each will possess it to a greater or less degree, depending on whether he values or disdains it. The responsibility lies with the one who makes the choice; the god has none. . . . The models of the lives were placed on the ground before them. There were far more of them than there were souls present, and they were of all kinds . . . There were tyrannies among them . . . There were lives of famous men, some . . . for . . . beauty . . . others for strength . . . others still for their high birth and the virtue or excellence of their ancestors. And there were also lives of men who weren't famous for any of these things. And the same for lives of women. But the arrangement of the soul was not included in the model because the soul is inevitably altered by the different lives it chooses. But all the other things were there, mixed with each other and with wealth, poverty, sickness, health, and the states intermediate to them. (617-618)

This (original) "original position" of the soul, of course, is a fiction: we choose our lives, so to speak, while living them, our souls already altered by the lives we have led. But it is a powerful image. It encourages us to adopt a perspective on our lives that treats them as objects to be judged as wholes, and to seek a kind of order in and understanding of them akin to that we seek elsewhere, and to attach to it our aspirations for our lives. Socrates remarks:

> [I]t seems that it is here . . . that a human being faces the greatest danger of all. And because of this, each of us must neglect all other subjects and be most concerned to seek out and learn those that will enable him to distinguish the good life from the bad and always to make the best choice possible in every situation. He should think over all the things we have mentioned and how they jointly and severally determine what the virtuous life is like. That way he will know what the good and bad effects of beauty are when it is mixed with wealth, poverty, and a particular state of the soul. He will know the effects of high or low birth, private life or ruling office, physical strength or weakness, ease or difficulty in learning, and all the things that are either naturally part of the soul or are acquired, and he will know what they achieve when mixed with one another. (618-619)

The specialization attendant upon professionalization in philosophy, without which one cannot do philosophy seriously—for despite its detractors, philosophy has made enormous progress in the last 2600 years—means that few of us can really be professional philosophical polymaths. We tend to restrict at least our professional philosophical attention to a field or area or figure. This restriction of our professional efforts, of the greatest concentration we bring to bear on philosophical problems, alters our approach to philosophy, and compartmentalizes our thinking outside our professional lives as well. Philosophy becomes a mere profession for us and fails to inform our lives—it fails to have that practical import for us which falls out of its goal of synoptic understanding; we fail to examine our lives in examining the questions of our professional interest.

This is not a philosophical thesis. It is a psychological thesis. And I do not say that there are not exceptions. But given that we are creatures of finite resources, of finite time and energy, of finite intellect, and given the encouragement of the institutional arrangements we choose to live within to pursue philosophy, it takes a kind of constant effort not to find that our lives have been left out of our philosophy. The danger of leaving one's life out of one's philosophy threatens, I think, even those among us whose professional interests lie in questions about value, character, and ethical and political life, where we might think the barriers are easier to cross. Would it not be an irony if, in considering the models of lives available for us, we chose the life of the professional philosopher only to find that it leads more often than not away from that synoptic understanding which was our motivation in choosing it?

I said at the outset that I would not try to answer the question that I would raise. You may be wondering what exactly the question is. I said it was a practical question. The question is simple. It is this: How can we retain our allegiance to philosophy while being professional philosophers? It sounds odd, stated in that way, but I hope to have at least indicated why I think there is a kind of special problem about it that is worth reflection.

I don't have an answer to it. I think it is the sort of question the possibility of an answer to which is best exemplified by examples of lives that are successful responses to the problem it raises. I doubt my own life has been like that though. And, disappointingly, like all practical questions, it shows a certain imperviousness to philosophical reflection, so that where we would most like to apply philosophy it is most likely to fail us.

Since one must end somewhere before a topic (and one's audience) is exhausted, I will end here. And that is enough, in any case, for one evening. Let me end on an optimistic note, following the precedent set by last year's president and setting the stage for the next:

Long enough your attention I have held
With relief I pass the torch to Schönfeld.

Aristotle and Supervenience Physicalism

Graduate Essay Prize Winning Paper
of the 47th Annual Meeting of the
Florida Philosophical Association

Jeremy Kirby, *Florida State University*

Introduction

In an article entitled "Is an Aristotelian Philosophy of Mind Still Credible? A Draft," Myles Burnyeat suggested that we might do "what the seventeenth century did . . . [with the Aristotelian concept of the mind] . . . junk it."[1] Burnyeat buttressed this controversial claim, in large part, on the premise that it is difficult to believe that mental facts are not supervenient on physical facts in the wake of post-enlightenment thinking.[2] Various valiant attempts to save Aristotle's philosophy of mind from being junked soon followed. One strategy that found favor among some scholars was that of arguing that Aristotle's physics really is not in conflict with the idea that mental facts supervene upon physical facts. Scholars such as Michael Wedin and Victor Caston read Aristotle as maintaining a supervenience thesis in *Physics* 7.3.[3]

I disagree with the view that ascribes supervenience physicalism to Aristotle. The general strategy for providing support for my view will run as follows: I will first aim to discredit the view that ascribes supervenience physicalism, hereafter (SV), to Aristotle on the basis of *Physics* 7.3. Thereafter, I will turn to more psychological and biological texts to argue that Aristotle's central views therein are unfriendly to (SV).[4]

Physics 7

In his preface to the Loeb edition of the *Physics*, Vol. 2, a revision of Philip Wicksteed's translation, F.M. Cornford reports that it was Wicksteed's opinion that Book VII was not the product of Aristotle but the product of an "acute and competent Aristotelian."[5] In the preface to Book VII itself, Cornford indicates the following:

> Simplicius, in his introduction to this Book, remarks that the more important and relevant problems treated in it are discussed in more detail in Book VIII. Some ancient critics accordingly regarded Book VII as superfluous, and Eudemus passed it over. Themestius treats it in summary fashion. Simplicius himself conjectures that

> Aristotle wrote Book VII at some earlier time and, when he had dealt with some of its topics more fully in Book VIII, allowed it to stand as a sort of introductory study.[6]

In a similar vein, W.D. Ross, in his commentary on the *Physics*, maintains that " . . . here are several indications that the book is not an integral part of the *Physics*, but is, even if it be by Aristotle, an excrescence of the main plan . . ."[7] A legitimate impression to take from these remarks is that there is some reason for the suggestion that Book VII is not an ideal representation of Aristotle's mature view—whatever that may be. And given such an impression, it seems *prima facie* surprising that supporters of (SV) have sought to exonerate Aristotle from Burnyeat's criticisms by appealing to such a text.

In any case, in broad outline of Book VII, Aristotle argues in the first chapter for the claim that whatever is moved is moved by another. In the second chapter, he argues that movement and the moved are always together. In the third chapter, Aristotle elaborates on one of his claims made in the second chapter, namely, that all alteration pertains to sensible qualities. In the penultimate chapter, Aristotle provides a comparison of movements. The final chapter discusses the proportion of movements.

***Physics* 7.3**

It is worth noting that whether Aristotle endorses (SV) or not in chapter three, his primary goal in the third chapter is not that of establishing (SV). His primary goal is, rather, that of arguing for the claim that all alteration pertains to sensible qualities. In fact, I think it is fair to say that his proposal for satisfying that goal is rather unsatisfying. Cornford[8] and Ross[9] both rightly maintain that Aristotle does not argue directly for his conclusion, call it C1, that all alteration pertains to sensible qualities. Rather, he supports C1 by refuting what he takes to be the most putatively formidable counterexamples. These putative counterexamples are the cases of shapes and figures, on the one hand, and states of the body or soul on the other. Aristotle does not, as one might initially suspect, argue that shape and figure, and states of the body and soul, in fact pertain to the sensible. Rather, he argues that such things are not in fact alterations. What I take to be a relatively uncontroversial outline of the argument runs as follows.

1. The two cases most likely to be thought of as counterexamples to C1 are (a) shapes and figures or (b) states of the body or soul.
2. If shapes and figures were alterations, then the resultant would retain the name of the material.
3. When shapes are acquired, the resultant does not retain the name of the material.
4. Shapes and figures are not alterations.

5. If shapes and figures are not alterations, then shapes and figures cannot be counterexamples to C1.
6. Shapes and figures cannot be counterexamples to C1.
7. If something is a perfection or defect, then it is not an alteration.
8. States of the body or soul are perfections or defects.
9. States of the body or soul are not alterations.
10. If states of the body or soul are not alterations, then such states are not counterexamples to C1.
11. States of the body or soul are not counterexamples to C1.

Wedin's Interpretation of 7.3

Because Wedin has been the most recent and outspoken proponent of (SV), I will consider his argument to be representative of the view. He argues as follows. In the course of establishing that all alteration pertains to sensible qualities, Aristotle maintains that somatic states exist in virtue of a particular relation:

> And in like manner we regard beauty, strength, and all the other bodily excellences and defects. Each of them exists in virtue of a particular relation and puts that which possesses it in a good or bad condition with regard to its proper affections, I mean those influences that from the natural constitution of a thing tend to promote or destroy its existence.[10]

In support of C1, Aristotle seems to argue that somatic states are not alterations, they merely exist in virtue of a particular relation. Let this be premise (1) in the following segment of reasoning.

1. If x is a φ-state, then x exists in virtue of a particular relation.

Wedin thinks that one can assume, "what seems harmless," that if x exists in virtue of a particular relation, x is a relative.[11] So let this be premise (2):

2. If x exists in virtue of a particular relation, then x is a relative.

However, Aristotle seems relatively clear on the point that the states in question are not alterations, and, furthermore, there are not alterations, generations, nor changes of such states.

> Since then relatives are neither themselves alterations nor the subjects of alteration or of becoming or in fact any change whatever, it is evident that neither states nor the processes of losing or acquiring states are alterations. (246b10-15)

Thus, premise 3 runs:

3. If x is a relative, then (a) x is not itself an alteration and (b) there is no alteration, generation, or change of x.

Premises 1-3, of course, entail the conclusion that:

4. If x is a φ-state, then (a) x is not itself an alteration and (b) there is no alteration, generation, or change of x.

However, conclusion (4) presents a difficulty. For in the lines that follow Aristotle says:

> It is evident that neither states nor processes are alterations, though it may be true that their *becoming* (γιγνεσθαι) or *perishing* (φθειϱεσθαι) is necessarily, like the becoming or perishing of a specific character or form, the result of the alteration of certain other things, e.g., hot and cold and dry and wet elements, whatever they may be, on which states primarily depend.[12]

This issues in the following difficulties. First, Aristotle has just maintained that states of the body and soul are not generated, as this was the conclusion reached in premise (4) above. But just a few lines later he speaks of the generation (γιγνεσθαι) (and destruction) of such states. Whence comes the following apparent contradiction:

(Φ) Aristotle says the φ-states are generated and that φ-states are not generated.

Secondly, the reader finds Aristotle likening φ-states to the case of a "specific character or form" (ειδος). Thus, the generation of φ-states is like the generation of form. And the idea that forms are generated is in direct conflict with what Aristotle has to say in *Metaphysics* Book VII, chapter 8, namely that form is not generated. Hence, there is a second apparent contradiction:

(E) Aristotle says that forms are generated and that forms are not generated.

Wedin, however, has a proposal. *Metaphysics* Book VII, chapter 8, does seem to allow for accidental generation of form (ϰατα συμβεβεϰος).[13] Hence, the tension in **(E)** can be resolved by acknowledging that when, at 246b10-12, Aristotle speaks of the generation of form, his locution is elliptical for "generation of form by accident." And "generation of form by accident" need not mean the same thing as, one might say, "generation of form simpliciter," which is clearly unacceptable to Aristotle in *Metaphysics* Book VII, chapter 8. Of course, the same reasoning goes *mutatis mutandis* for **(Φ)**. Aristotle likens φ-states to forms, as Wedin sees it, for heuristic reasons:

> But the inclusion of form in [the] analogy serves a more important point . . . [than resolution of an apparent contradiction] . . . It stands as a clear case, introduced to explain the less familiar and more difficult case of generation of φ-states. On the clear case, the form is produced when certain matter is organized in a certain way, that is, when matter undergoes alterations of a certain sort. Parity of reasoning . . . would, therefore, lead us to expect that φ-states are also generated.[14]

Provided one thinks that Wedin's statements are intelligible here—as I will explain below I do not—one question that springs to mind concerns the nature of the "matter [that] undergoes alterations of a certain sort." If Wedin is right in claiming that Aristotle believes that form is produced when matter undergoes alterations of a certain sort, we should expect Aristotle to provide some examples of this process. And, as a matter of fact, we find Aristotle saying something that seems to fit into the picture Wedin wants to present. Recall what Aristotle says at 246b10-12:

> It is evident that neither states nor processes are alterations, though it may be true that their *becoming* (γιγνεσθαι) or *perishing* (φθειϱεσθαι) is necessarily, like the becoming or perishing of a specific character or form, the result of the alteration of certain other things, e.g., hot and cold and dry and wet elements, whatever they may be, on which states primarily depend.[15]

Admittedly, the language in this passage is not unfriendly to (SV). And, given the means by which the problems that arose vis-à-vis **(Φ)** and **(E)** were dispensed with, a counter-argument to the effect that forms and φ-states are not generated is, seemingly, not at our disposal. Moreover, the reader finds Aristotle claiming that forms and φ-states are "the result of alteration of certain other things, e.g., hot and cold and dry and wet elements." It is not a lengthy reach, therefore, to say that macrophysical and formal states supervene upon the micorophysical states Aristotle countenances, i.e., the hot and cold and dry and wet.

Accepting Wedin's proposal has ramifications for psychological and noetic states as well. For Aristotle maintains that these too are relatives. Hence, it seems that these states too will fit into the segment of reasoning I referred to above.[16] By analogy, therefore, psychological states and noetic states will be thought to supervene upon the microphysical.

Wedin's argument is complex. Here is a concise summary of the reasoning he expects of his reader: First, he points to two apparent contradictions in the text. His proposal for resolving these tensions is that form is "in some weak sense" produced according to accident.[17] Once it is admitted that form can be generated, he can run his argument:

1. Form is generated by accident (*Metaphysics* 7.8).
2. If form is generated by accident, form must come about solely by material-efficient means, i.e., not by formal/final means.
3. Form must come about solely by material-efficient means (1,2).
4. If form must come about solely by material-efficient means, alteration at the microphysical level seems like the best candidate for the production of form.
5. Alteration at the microphysical level seems like the best candidate for the production of form (3,4).

Furthermore, it seems that this reasoning goes *mutatis mutandis* for φ-states, psychological states, and noetic states. One can substitute, it seems, any of these terms for form. Aristotle says that φ-states are relatives, form is used to explain the situation with such relatives, and he goes on to say that noetic states[18] and psychological states[19] are relatives like φ-states.

Is Wedin's Proposal Acceptable?

In this section I want to address what seems to be a significant circularity in Wedin's language, in order to object to premise (2) in the argument just given. Consider again what Wedin has said about the analogy drawn between form and physical states.

> But the inclusion of form in [the] analogy serves a more important point . . . [than resolution of an apparent contradictions] . . . It stands as a clear case, introduced to explain the less familiar and more difficult case of generation of φ-states. *On the clear case, the form is produced when certain matter is organized in a certain way, that is, when matter undergoes alterations of a certain sort.* Parity of reasoning . . . would, therefore, lead us to expect that φ-states are also generated.[20]

Recall that Wedin's project is to make palatable the idea that Aristotle accepted supervenience physicalism. This means making palatable the idea that Aristotle would accept the view that explanation could, in theory, rely solely on material-efficient causation.

Consider in isolation Wedin's claim: "On the clear case, the form is produced when certain matter is organized in a certain way, that is, when matter undergoes alterations of a certain sort."[21] I take it that Wedin thinks that forms supervene upon "certain material." But notice that the matter upon which form supervenes is "organized in a certain way." What can that which is "organized in a certain way" be if it is not the form of matter? The alterations and matter upon which Aristotle is thought to have form rely are of a "certain sort." The matter, therefore, which serves as subvenient is, to some extent, informed. Hence, the picture will be something like this: F is produced by material m & form f. Is f produced by alterations of matter of a certain kind, i.e., f & m? If this is the case, and it seems that Wedin's interpretation is committed to such a picture, there will be forms all the way down. If there is a lowest subvenient domain, there will be matter of a certain sort, i.e., matter that is informed, if it is the kind of matter that can be responsible for its supervenient counterpart. And if this is the case, form will not, therefore, be eliminable at the lowest level at which a target property is said to supervene upon a base property. Therefore, there will be either at least one form that is not generated by certain alterations of certain matter or there will be an infinite regress of subvenient levels. This presents the reason for thinking premise (2) is false.

If form is generated by accident, form must come about solely by material-efficient means. The material-efficient means, according to Wedin, will be "certain matter" undergoing "alterations of a certain sort". But that which is "certain matter" has form. So to accept (2) in this way is to accept a falsehood or an infinite regress. The former, needless to say, is unattractive. The latter is not Aristotelian. This is perhaps reason enough to regard (SV) a lost cause.

I will leave aside complex and controversial issues concerning the existence of *materia prima.* But I do not think doing so presents a deficiency on my part. Either *materia prima* is of a certain sort or it is not. If it is a certain sort, i.e., has a form, form at the lowest level will be uneliminable. If it is not matter of a certain sort, it is not the kind of thing upon which, on Wedin's reasoning, other states may supervene.

Another Argument Against Deduction From the Bottom Up

Scholars have long recognized that Aristotle's views seem unfriendly, indeed hostile, to projects aimed at unifying the sciences such as reduction.[22] Teleology, in some form or another, is the virtue of Aristotle's project from the nonreductionist's perspective. It is the fly in the ointment from the reductionist's point of view. In this section, I provide an argument against (SV) that relies on some of Aristotle's teleological views. The argument to be considered should be prefaced by recalling some citations that illustrate Aristotle's teleological commitments. For example, in *Physics*, 200a7-11, Aristotle states the following:

> Similarly in all other things which involve production for an end; the product cannot come to be without things which have a necessary nature, but it is not due to these [*δια ταυτα*] (except as its material (αλλ' η ὡς υλην)); it comes to be for an end (αλλ' ενεκα του).

No doubt, Aristotle countenances material necessity in this passage. But material necessity is given a subordinate role. Material comes to be for the final cause. And this subordinate relation of material necessity to final cause is a recurring theme for Aristotle. Consider *Generation of Animals* 5.8 789a8-b8:

> Once [the front teeth] are formed, they quickly fall out on the one hand for the sake of the better, because what is sharp quickly gets blunted, so that [the animal] must get other new ones to do the work of [tearing off food]; . . . on the other hand they fall out from necessity, because the roots of the front teeth are in the thin part [of the jaw], so that they are weak and easily work loose. . . Democritus, however, neglecting to mention that for the sake of which things [happen in the course of nature], refers to necessity all the things that nature uses—things are indeed necessitated in that way, but that does not mean they are not for the sake of something, and for the sake of what is better in each case. So nothing prevents [the front teeth] from . . . falling out in the way he says, but it is not on account of these factors (*δια ταυτα*) that they do, but on account of the end (*δια τελος*): they are causes as sources of motion and instruments of matter.[23]

In this passage, the reader finds Aristotle explicitly recognizing material necessity—a material necessity that he characterizes as Democritean. Still, material necessity is clearly considered subordinate to final cause. Both the teleological aspect and the material aspect are aspects of *rerum* *natura.*[24] But the material aspects seem to be at the invitation, moreover, of the teleological. At Physics 2.7, the matter *comes to be* for the sake of the end. And in *Generation of Animals* 5.8, the thing comes about *δια τελος* (because of the end) rather than *δια ταυτα* (because of these things).

In light of these considerations, it is interesting that Aristotle thinks that final cause, efficient cause, and formal cause, are often the same thing:

> (A1) Now the causes being four, it is the business of the physicist to know about them all . . . the matter, the form, the mover, [and the] 'that for the sake of which.' The last three often coincide.[25]

What is more, we read in the *De Anima* that soul is a paradigmatic example of such a coincidence:

> (A2) Soul is the cause of the living body in the three ways we have distinguished . . . (a) mover . . . (b) the end . . . (c) the essence of the whole living body.[26]

It is, therefore, abundantly clear that (SV) is vulnerable to an argument by *reductio ad absurdum*:

1. Soul = form (*De Anima* 412a20).
2. Soul = efficient cause of the living body (A1 & A2).
3. Form is the efficient cause of the living body (1,2).
4. Form is generated when material and efficient causes generate a particular substance independently of the form (assuming (SV)).
5. A living thing is a particular composite substance.
6. A living thing is a particular composite substance whose form is generated by an *efficient* cause and material cause independently of form (4,5).
7. A living thing is a particular composite substance whose form is generated by a *formal* and material cause independently of form (3,6).

Obviously, to say that form is generated by form and that form is generated independently of form—that is, not generated by form—is contradictory. (SV) is false.

There is a temptation here to say that Aristotle's description of formal, efficient, and final causes as "coinciding" does not necessarily entail that Aristotle thought, in the case of soul, that formal=final=efficient. One might accept the idea that when Aristotle describes causes as "coinciding," he means to indicate that they are "temporally and spatially contiguous." But form, presumably, is not a spatio-temporal entity, so this can be ruled out. One might instead maintain that Aristotle means "copresent," in some non-spatial way, whatever that means. But to accept this would be a mistake. For Aristotle's treatment of soul applies to living things. Living things are found in the sublunary realm. Entities belonging to the sublunary realm have material aspects. And material aspects, as well, will be "copresent" in sublunary substances. Aristotle does not say that the last four, i.e., all *aitiai* coincide. He says the "last three coincide." To say that the last three coincide, where "coincidence" means "copresence," when all four *aitia* are always copresent, is implausibly pleonastic. Aristotle means by "coincidence," in this context, "identical."

An Alternative Way of Rendering Consistency

To leave things thus would, however, visit the two previously mentioned contradictions on Aristotle:

(Φ) Aristotle says that φ-states are generated and that φ-states are not generated.

(E) Aristotle says that forms are generated and that forms are not generated.

Can Wedin's solution to these difficulties be accepted without accepting (SV)? (SV), as we have seen, leads to manifold difficulties. A resolution of these apparent contradictions that does not involve acceptance of the idea of form being generated, even in some "weak sense," would therefore be preferable. Recall the supervenient-friendly text:

> It is evident that neither states nor processes are alterations, though it may be true that their *becoming* (γιγνεσθαι) or *perishing* (φθειϱεσθαι) is necessarily, like the becoming or perishing of a specific character or form, the result of the alteration of certain other things, e.g., hot and cold and dry and wet elements, whatever they may be, on which states primarily depend.[27]

If it is permissible to say that form is, in some ontological sense, generated, then the above statement appears, admittedly, supervenient friendly. I have argued, heretofore, that (SV) is in direct conflict with several of Aristotle's theoretical commitments and should therefore be rejected. Yet one might feel the pull of (SV) in connection with this passage and the resolution of the two apparent contradictions **(Φ)** and **(E)**.

Fortunately, I think there is a better resolution than that which Wedin offers. Consider, for the moment, the following alternate translation from Cornford and Wicksteed:

> It is clear that neither are habits (εξεις) such, not the acquisition or loss of them [i.e., alterations]; though it might be that, just as with the characteristics or forms we have already spoken of, the *formation* [γιγνεσθαι] and destruction [φθειϱεσθαι] of habits may involve the modifications of certain factors, (say) the heat or cold or dryness or moisture of the physical elements, or the proper seats of the habits whatever they may be.[28]

This translation does, no doubt, take some liberties. And a more literal translation could indeed be given. But the salient point is that γιγνομαι can simply mean "happen." Aristotle need not be saying that forms are generated. Aristotle, it seems, can simply be saying that forms come to be or happen in the sense of being instantiated. This resolves the apparent difficulty that **(E)** is thought to present. And this is not incompatible with his principle in *Metaphysics* 7.8 which, in effect, says that forms are not generated in the sense of being produced or born.

At this point, one might be inclined to accept that γιγνομαι is not always used to denote genetic change and still reject the present line of reasoning on the grounds that the same line of reasoning does not apply to φθειϱεσθαι. After all, in the passage, φθειϱεσθαι is used in connection with γιγνεσθαι, which is Aristotle's antonym for genesis. However, H. Bonitz indicates that φθειϱειν has for a synonym διαλυεσθαι.[29] And διαλυω is a word that usually means "to dissolve" or "part." This meaning dovetails nicely with the above interpretation of γιγνεσθαι (most generally, "comes-to-be") where γιγνεσθαι is taken in the non-genetic sense. The form need not be undergoing destruction. It is reasonable to assume that Aristotle intended only to say that forms part way with the particulars that instantiate them.

Needless to say, one will need to apply similar reasoning to the case of physical states. Can this be done? What does it mean to instantiate a physical state? This is not, however, a major difficulty, if "matter" is treated as a relative term. And treating "matter" as a relative term has proven to be a useful way of reading Aristotle.[30] That which is matter for one thing, e.g., the bronze of the statue, is form for another, e.g., the elements of the bronze. Hence, it makes sense as well to talk of physical states as instantiated. Indeed, according to R.D. Hicks, Aristotle does occasionally use the "state" (*hexis*) as a synonym for form.[31]

Still, one might ask what Aristotle means by generation ϰατα σμβεβεϰος (by accident) in *Metaphysics* 7.8. I think that generation according to accident can be best viewed as a linguistic phenomenon. For example, Aristotle says that the craftsman brings the circle into the matter (presumably, to make a shield). Hence, we say that the craftsman produces the shield. The shield is a circle. Is the relation transitive? Does the craftsman produce the circle? Aristotle answers yes ϰατα συμβεβεϰος. But there can be little doubt from the context that Aristotle is trying to answer "no" to this question. The thesis of the chapter is that form is not generated. What I find interesting is that Aristotle, in *Physics* 7.3, seems to suggest a way out of this difficulty. Recall that in that chapter Aristotle offered the following argument:

1. The two cases most likely to be thought of as counterexamples to C1 are (a) shapes and figures or (b) states of the body or soul.
2. If shapes and figures were alterations, the resultant would retain the name of the material.
3. When shapes are acquired, the resultant does not retain the name of the material.
4. Shapes and figures are not alterations.

Moreover, if shapes and figures are not alterations, shapes and figures cannot be counterexamples to C1.

One example that Aristotle provides, at 245b9-14, in support of premise (3) runs as follows:

> In the first place, when a particular formation of a thing is completed, we do not call it by the name of its material: e.g., we do not call the statue 'bronze' or the pyramid 'wax' or the bed 'wood,' . . . but we use a derived expression and call them 'brazen,' 'waxen,' or 'wooden,' 'respectively'.[32]

Thus, in answer to the question, "Does the craftsman make the circle in making the shield?" Aristotle can treat the problem as a mere linguistic difficulty. As things are in *rerum natura*, the craftsman does not make the shield a circle, but makes it circular. According to ordinary language, sometimes we are inclined to say that since the circle belongs to the shield and the shield is produced by the craftsman, then the circle is "produced" by the craftsman. But this is not to say that a more precise paraphrase is not available. A more precise articulation is to say not that the shield is a circle but that the shield is circular. So there are means more benign, I submit, for dealing with the apparent contradictions than those that Wedin pursues in his defense of (SV).

Notes

[1] M. Burnyeat, "Is an Aristotelian Philosophy of Mind Still Credible? A Draft," *Essays on Aristotle's De Anima,* eds. Martha Nussbaum and Amelie Rorty (New York: Oxford UP, 1999): 26.

[2] Burnyeat 23.

[3] Michael Wedin, "Content and Cause in Aristotelian Mind," *Southern Journal of Philosophy* 31, Supplement (1992) 49-105; Victor Caston, "Aristotle and Supervenience," *Southern Journal of Philosophy* 31, Supplement (1992):107-135. In line with Jaegwon Kim's (1984) distinction between strong and weak supervenience (153-176), Wedin reads Aristotle's putative supervenience as that of the strong variety. Caston argues that Aristotle only allowed for the weak variety of supervenience. However, Caston subsequently recants and reads Aristotle as maintaining the more robust strong supervenience in "Epiphenomenalisms, Ancient and Modern," *The Philosophical Review* 106, (1997): 309-363.

[4] Burnyeat, I believe, overreacts in suggesting that Aristotle's concept of the mind might be junked. Supervenience does not seem to enjoy the explanatory status that it once held. I will not, however, in this paper, argue that Burnyeat is wrong.

[5] F.M. Cornford and Phillip Wicksteed, "Introduction," *Aristotle: The Physics* (Cambridge, Mass: Harvard UP,1929): v.

[6] F.M. Cornford and Phillip Wicksteed, Commentary, *Aristotle: The Physics* (Cambridge, Mass: Harvard UP, 1929): 204.

[7] W.D. Ross, *Aristotle's The Physics: Text with Commentary* (Oxford: Oxford UP, 1955): 15.

[8] Cornford and Wicksteed, Commentary, 228.

[9] Ross 674.

[10] Aristotle, *Physics* 7.3 246b5-10, ed. Richard McKeon, *The Basic Works of Aristotle* (New York: Random House, 1968). All subsequent references to Aristotle are from the McKeon edition unless otherwise noted.

[11] Wedin 53. Indeed, I think this is a harmless assumption because Aristotle seems to say as much in the seventh chapter of the *Categories*.

[12]Aristotle, 246b10-12. My italics.

[13]Aristotle, *Metaphysics* 1033a30.

[14] Wedin 54.

[15] Aristotle, *Physics*, ed. Richard McKeon (New York: Random House, 1968): 246b10-12.

[16] Aristotle, *Metaphysics* 247b1-9; 247a4-7, respectively. Cf. premises 1-4 on pp. 5-6.

[17] Wedin 54.

[18]Aristotle, *Metaphysics* 247b1-9.

[19]Aristotle, *Metaphysics* 247a4-7.

[20]Wedin 54. My italics.

[21]Wedin 54.

[22] Jaegwon Kim has argued effectively that strong supervenience is nagel-reduction. *Supervenience and Mind* (Cambridge: Cambridge UP, 1993): 53.

[23] Tr. John Cooper "Hypothetical Necessity," *Philosophical Issues in Aristotle's Biology*, eds. A. Gotthelf and J. Lennox (Cambridge: Cambridge UP, 1987): 258.

[24] Aristotle, *Physics*, 246b10-12. Here, I am following Julius Moravcsik (1991), John Cooper, (1987), William Wians (1992), and others, in rejecting Martha Nussbaum's (1983) position.

[25] Aristotle, *Physics* 2.7 198b 25ff.

[26] Aristotle, *De Anima* 415b8-12.

[27]Aristotle, *Physics,* 246b12-18. The last clause of this section could be thought of as suggesting a dependency relation supportive of (SV). The Greek, I think, is ambiguous. Compare Ross's note on the line: "or whatever it may be in which the states directly reside."

[28] Aristotle, *Physics,* trans. F.M. Cornford and P.H. Wicksteed (Cambridge, Mass: Harvard UP, 1929): 198b 25ff.

[29] H. Bonitz, *Index Aristotelicus* (Graz: Academishe Druck- u. Verlagsanstalt, 1955): 816, 55.

[30] See Jonathon Lear, *The Desire to Understand* (Cambridge: Cambridge UP, 1999) sections 2.1, 2.2, and 2.4.

[31] R.D. Hicks, *Aristotle: de Anima* (Oxford: Oxford UP, 1907): 501. He cites *Metaphysics* 12, 107a11, 1069b34, 1070b11; 8, 1044b32.

Works Cited

Aristotle. *Aristotle: de Anima.* Trans. and Commentary R.D. Hicks. Oxford UP, 1907.

Aristotle. *Aristotle: The Physics.* Trans. and Commentary F.M. Cornford and P.H. Wicksteed. Cambridge: Harvard UP, 1929.

Burnyeat, M. "Is an Aristotelian Philosophy of Mind Still Credible? A Draft." *Essays on Aristotle's De Anima.* Eds. Martha Nussbaum and Amelie Rorty. New York: Oxford UP, 1995. 15-26.

Bonitz, H. *Index Aristotelicus.* Graz: Academishe Druck- u. Verlagsanstalt, 1955.

Caston, Victor. "Aristotle and Supervenience." *Southern Journal of Philosophy,* 31 Supplement (1992): 107-136.

Caston, Victor. "Epiphenomenalisms, Ancient and Modern." *The Philosophical Review* 106 (1997): 309-363.

Cooper, John. "Hypothetical Necessity." *Philosophical Issues in Aristotle's Biology.* Eds. A. Gotthelf and J. Lennox. Cambridge: Cambridge UP, 1987. 243-274.

Cornford, F.M. and P.H. Wicksteed. "Introduction." *Aristotle The Physics.* Cambridge: Harvard UP, 1929.

Hicks. R.D. *Aristotle: de Anima.* Oxford: Oxford UP, 1907.

Kim, Jaegwon. *Supervenience and Mind.* Cambridge: Cambridge UP, 1993.

Lear, Jonathon. *The Desire to Understand.* Cambridge: Cambridge UP, 1999.

McKeon, Richard, ed. *The Basic Works of Aristotle.* New York: Random House, 1968.

Moravcsik, Julius. "What Makes Reality Intelligible? Reflections on Aristotle's Theory of *Aitia.*" *Aristotle's Physics: A Collection of Essays.* Ed. Lindsay Judson. New York: Clarendon, 1991. 31-48.

Nussbaum, Martha. "Saving Aristotle's Appearances." *Language and Logos: Studies in Ancient Philosophy Presented to G.E.L. Owen*. Cambridge: Cambridge UP, 1983. 267-93.

Ross, W.D. *Aristotle's The Physics, Text with Commentary*. Oxford: Oxford UP, 1955.

Wedin, Michael. "Content and Cause in Aristotelian Mind." *Southern Journal of Philosophy,* 31 Supplement (1992): 49-105.

Wians, William. "Saving Aristotle from Nussbaum's *Phainomena*." *Essays in Greek Philosophy V: Aristotle's Ontology*. Eds. Anthony Preus and John P. Anton. Albany: SUNY Press, 1992. 133-150.

Hempel on Intertheoretic Reduction

Winner of the Gerrit and Edith Schipper Undergraduate Award
for Outstanding Undergraduate Paper at the
47th annual meeting of the Florida Philosophical Association

David Barnett*, New College of Florida*

Introduction

The question of whether all living things are really just complex physical ones, or whether instead there are biological entities or characteristics that cannot be fully characterized in physical terms, has historical roots buried centuries deep. Carl Hempel considers this question as an empirical one for modern science to address.[1] Hempel's concern in this paper is not with the answer to the question, but rather with the methods by which it may be evaluated. He considers the position of those he calls "mechanists," that all living things and their biological characteristics are nothing more than complex physical systems, as equivalent to the view that in some significant sense all accurate biological theories are implied by physical ones.[2] In doing so, Hempel seeks to draw conclusions regarding the unity of science more broadly. I will argue that Hempel's account, though perhaps succeeding in a crucial first step, fails on numerous points afterwards. Using the morals that may be drawn from these failures, I suggest rough outlines of some alternative accounts of intertheoretic reduction.

The Project Outlined: Reduction of Theories

Contrasted with vitalism, the position of the mechanists, says Hempel, amounts to the following two claims:

> (M1) All the characteristics of living organisms are physico-chemical characteristics—they can be fully described in terms of the concepts of physics and chemistry.[3]
>
> (M2) All aspects of the behavior of living organisms that can be explained, can be explained by means of physico-chemical laws and theories.[4]

This formulation of the mechanist position is never explicitly argued for, but I will not here draw it into question. Perhaps there are other conditions that must be the case in order for the mechanist position to be correct, or perhaps it might be that living things are nothing more than physical ones, although one or both of these conditions are not satisfied. Here, however, I will assume that Hempel has got this much right about what it is for a biological entity to be merely a physical one. Moreover, I will grant Hempel that generalized versions of M1 and M2 are both necessary and sufficient conditions for an analogous identity between, say, mental entities and physical ones, or sociological systems and complex systems of psychological entities.

Hempel further interprets M1 and M2 in terms of the relationship between biological and physical theories. M1, he claims, requires that the terms of biology be extensionally definable in physical terms.[5] Contrasted with intensional definition, this sort of definition requires only that the biological term being defined share its extension with some physical term. He sees M2 as requiring that all biological facts, including all the laws of accurate biological theories, be deducible from the laws of physics.[6]

The question whether biological systems are nothing more than physical ones is treated by Hempel as equivalent to the question whether biology is reducible to physics. The terms "biology" and "physics" refer to theories—linguistic entities that describe biological and physical things.

Equating mechanism with the view that biology is reducible to physics is accomplished by way of something analogous to the principle of semantic ascent. According to Quine, the principle of semantic ascent translates the claim that there are wombats in Australia as semantically equivalent to the claim that 'wombat' applies to something in Australia.[7] In so doing, it allows us to "ascend" from questions regarding *things* to questions regarding the *words* that refer to them. In an analogous manner, I believe, the conflation of mechanism with biological-physical reductionism requires us to treat the statement "biological properties are nothing more than physical properties" as equivalent to the statement "biological predicates are in some sense reducible to physical predicates."

Using this analog to the principle of semantic ascent, we can convert M1 and M2 into the biological-physical reductionist claims R1 and R2:

(R1) The terms of biology are reducible to those of physics.

(R2) All laws of biology are reducible to those of physics.

R1 may be said to call for a reduction of terms and R2 for a reduction of laws.[8] As it is, "reducibility" in the case of terms can be defined trivially as the relationship which obtains between two sets of theoretical terms, A and B, such that A is reducible to B just in case all of the things to which A-terms refer are really just the things to which B-terms refer. The reduction of laws proceeds likewise.

It is the job, then, of an account of intertheoretic reduction to tell us, in a non-trivial way, what the reducibility relation involves. By Hempel's account, a reduction of terms amounts to extensional definition; deducibility is a necessary and sufficient condition for the reducibility of laws.

The Reduction of Terms

Let us now take a closer look at extensional definition as the mode by which a reduction of terms takes place. A biological term, by Hempel's account, can be defined by means of any physical term with the same extension. As Hempel explains things, the only relationship between a biological and a physical term necessary for a reduction is that which obtains between "human" and "featherless biped."[9]

This unusual view is based on a process of elimination. As Hempel sees it, extensional definition is our only plausible candidate for a mode of terminological reduction. He points out that it would be impossibly restrictive to expect the reduction of a particular biological term to follow analytically from the meaning of that term.[10] He fails to consider, however, any other possible modes of terminological reduction, never explaining why he considers these two possibilities exhaustive.

If an account of intertheoretic reduction is to be possible, then it is to be hoped that these two options—extensional definition and analytic equivalence—are not exhaustive, as neither of them work. As Hempel points out, analytic equivalence is not necessary for a terminological reduction.[11] It would be crazy to say that an equivalence with a physical term must be contained in the meaning of the biological term to be reduced for a reduction to be possible. However, extensional definition, which Hempel actually endorses, seems like an unnecessarily restrictive criterion for terminological reduction as well.

Hempel himself notes that the adoption of a particular terminological reduction often involves changing the intension as well as the extension of the term being reduced. He gives, as an example, the word "testosterone," which he says was originally defined as a male sex hormone produced by the testes. Once the term was reduced, the intension changed to a physico-chemical characterization, thereby widening the extension to include synthetic substances of the same chemical structure.[12] What Hempel fails to note is that, if the extension of "testosterone" was altered to include these substances, then its previous extension was not identical to the physico-chemical description to which it was reduced. Since the physico-chemical description to which testosterone was reduced includes synthetic substances under its extension and "testosterone" did not include these substances initially, then the two terms were not co-extensional.

One can try to avoid this dilemma by suggesting various *ad hoc* modifications to extensional definition. For instance, one could require, rather than complete extensional equivalence of the terms between which the reduction is taking place, merely a substantial overlap of some sort. But, it cannot be a matter merely of having nearly identical extensions. First, there are the obvious difficulties that mass terms like "testosterone" and "gold" present to such a story (how might one compare the degree to which their extensions overlap with corresponding physical terms, without the presence of discrete items in their extensions?). Second, it seems to make no difference, for example, whether synthetic testosterone exists in tiny or enormous quantities. In other words, the quantity of synthetic material that falls under the extension of the physico-chemical description in question but not under the extension of "male sex hormone produced by the testes," seems to make no difference in determining whether a reduction is appropriate.

A direction which seems much more plausible is to deny that extensions do indeed change after a reduction, thereby allowing one to maintain Hempel's overall account of terminological reduction. The extension of "testosterone," then, has always included chemically similar synthetic substances, even before its chemical structure was ever examined. As the standard definition of "testosterone" currently does include a physico-chemical characterization not included prior to reduction, one must either 1) claim that the definition of the biological term does not in fact determine its extension, or 2) grant that the definition determines the extension, although it never alters in such a way as to change the reduced term's extension.

Since synthetic substances are within the extension of "testosterone" now, yet clearly would not be included under the definition "male sex hormone produced by the testes," the apparent way to go is with option 1. A causal account of reference might be suitable for this task. One could claim, for instance, that once the reference of "testosterone" is fixed to some batch of testosterone, then the term will refer to all other instances of the same natural kind. Unfortunately, this view, coupled with Hempel's account of terminological reduction, makes it difficult to see how one might ever know whether a reduction is appropriate. If terminological reduction requires that the extensions of the terms in question match up, one needs a way of determining the extensions of those terms. If definitions do determine reference, then we can investigate the matter by checking to see whether all male sex hormones produced by the testes fall under the physical description being considered for reduction. Otherwise, it is not obvious how one might go about determining if a biological and a physical term had the same extensions, as the extensions themselves are unknown. This is not to say, of course, that causal theories of reference for natural kind terms imply that we can never know the extension of a term. Clearly, we know the extension of "water," and can know whether a particular liquid is water or not by testing to see if it is H_2O, as opposed to XYZ. However, this is a determination which we can make *after* the reduction has taken place. Before "water" was reduced to "H_2O," it is not obvious how we might have determined whether these terms were co-extensional.

The option that I see as more promising is 2. If the definition of "testosterone" really does denote all of the same things as the physical term to which it is being reduced, then a change of definition to include this physical term will not alter the extension of "testosterone." What has been taken for granted until this point is that the definition of "testosterone" as "a male sex hormone produced by the testes," was in fact the one in use prior to its reduction. Suppose that chemists have just provided a chemical description that apparently characterizes all instances of testosterone. They have also found that a particular synthetic substance, which has been produced in a laboratory as a byproduct of their experiments for years, falls under the same chemical characterization. It seems to me as though a terminological reduction would be appropriate just in case the synthetic substance had the same effects as natural testosterone if released into the male bloodstream. If the synthetic substance failed to promote growth in men in the same manner in which natural testosterone does so, we would say that the chemical characterization fails to capture something important about what it is for a substance to be testosterone. Conversely, if the synthetic substance shared all of these biological properties with natural testosterone, a reduction of the term would be in order. Despite the fact that "male sex hormone produced by the testes" is what one will find under the heading "testosterone" in the glossary of a pre-reduction biology textbook, it might be that the biological characteristics which in fact determine the extension of "testosterone" are associated with the effects of this substance on the male body.

Whether or not Hempel's account of terminological reduction can in this manner be defended against the charge of stringency (i.e., that it rules out legitimate cases of reduction), I will argue that it is clearly guilty of its converse, permissiveness. That extensional definition is insufficient for reduction should be apparent from the fact that many biological terms (especially those with very limited extensions) may have multiple unique physical descriptions with which they are co-extensional. In the cases in which these physical descriptions are theoretically equivalent (e.g., in the sense that having a particular net charge is equivalent to producing a certain electric field), this seems perfectly unproblematic. But in instances in which the physical properties in question are not theoretically equivalent, there will be multiple independent but equally correct reductions of the same biological term.

That it is possible for a particular biological term to be co-extensional with multiple physical terms, I take it, is *prima facie* quite plausible. Additionally, I will argue, this, in fact, is the case quite often. It is not difficult to come up with bizarre disjunctive physical properties that ought to be co-extensional with any particular biological term. This is especially easy in cases in which the biological term has only one or several items in its extension. For example, for a biological term that has only one object in its extension, any physical property that is possessed uniquely by that object will be co-extensional with the biological term. Were Hempel's account of terminological reduction correct, the biological term in question would be reducible to any physical property (however bizarre and disjunctive) that applied uniquely to this object.

This is not the case merely with terms with limited extensions. Take, for instance, the common example of the identification of water with the chemical compound H_2O. By Hempel's account, a reduction such as this is appropriate if and only if "x is H_2O" is co-extensional with "x is water." But similarly, "x is water" is also co-extensional with "x is H_2O or Li_3," as there are, presumably, no instances of trilithium anywhere.

One might make an *ad hoc* requirement that if we are to reduce some biological or descriptive term to a disjunctive physico-chemical description, then each of the disjuncts of the physico-chemical description must have something in their own extensions. Minor modifications like these, I think, must ultimately fail. The problems with extensional definition as a criterion for terminological reduction run much deeper than this.

Let's assume that there is some physical description—call it "ORGANISM"—to which the biological term "organism" is reducible. In order for there to be a complete reduction of the terms of biology to those of physics, it would also need to be the case that terms referring to particular species of organisms also be reducible to physical terms. In other words, just because "organism" has some physical reduction, it does not follow trivially that "bumble bee," "oak tree" and "turkey" all have physical reductions. By Hempel's account, a species term like "turkey" is reducible to a physical description if and only if that description applies to all turkeys and only to turkeys.

A particular turkey might have a mass of, say, 4.56968360857 kg at time t_1.[13] It is plausible to suppose that at the exact instant at which the turkey has this mass, this particular turkey will be the only organism to have precisely this mass. (If this does not seem equally plausible to the reader, substitute for a turkey some organism with an abnormally large mass, such as a blue whale.) Indeed, if we are specifying mass to the nearest ten nanograms, as above, it is probably true of any turkey with a mass m_i that it is the only organism at time t_i with that mass. The following physical property, then, is co-extensional with "x is a turkey":

> x is ORGANISM AND [x has mass m_1 at time t_1 OR x has mass m_2 at time t_2 OR . . . x has mass m_n at time t_n] where $m_{1.}$ . . . m_n specify the mass of each turkey that will ever live at times t_1 . . . t_n

Yet it is counter-intuitive that "turkey" is reducible to this physical description, which we may call "TURKEY." What seems to be delinquent about TURKEY is that, while it is co-extensional with "turkey," it fails to capture any of the salient biological features of turkeys. As in the case of testosterone, we would want to reduce the biological term only to a physical term that captures the biological properties that we most closely associate with that type of biological entity. In the case of testosterone, this means, roughly, that if we took any chemical of the given description and placed it in the bloodstream of a man, it would have the same effects as natural testosterone.

Similarly, while there are not in fact any physical objects which are in the extension of TURKEY which do not posses the biological properties of turkeys, it is possible, say, that a physical object with the biological properties of a dog was an organism with a mass of 4.56968360857 kg at time t_1. What is needed is a physical description that, as a matter of physical *necessity*, will have to apply to all and only those things with the biological properties exclusive to turkeys.

This matter is complicated by the fact that whether a particular physical description necessitates the possession of certain biological properties depends on the reductions which can be given for those biological properties. This is quite evident, again, in the case of testosterone. Whether a particular chemical structure will necessarily have the biological property of, say, causing growth in male humans, depends on the particular physico-chemical reductions that can be given for "growth," "male," and "human." If, for instance, the physico-chemical structure of the human body were drastically different from the way it actually is, then the chemical compound with which testosterone is actually identical would not have the biological properties that it actually possesses. There are probably numerous ways in which the chemical structure of organisms could be different (at least in an epistemic sense of "could") which would yield the same biological laws and observations. For this reason, the reduction of any biological term to a physical term depends on the physical reduction of other biological terms. A reduction of terms is an all or nothing affair—terminological reductions cannot be considered on a term by term basis.

As an illustration of this conclusion, let us consider the reduction of biological terms that are the result of the combination of other biological terms. For instance, "female mammal," "primate with no tail," and "purple swan." It is intuitively obvious that a reduction of one of these terms should be physically equivalent to the logical combination of the physical properties to which the individual terms can be reduced. If "purple" is reducible to the physical description PURPLE, and "swan" to the physical description SWAN, then "purple swan" is reducible to the physical description [PURPLE AND SWAN]. However, as has been argued previously, there will often be multiple physical properties with which a particular biological term is co-extensional. This is particularly obvious in the case of "purple swan," which is co-extensional with every physical description with nothing in its extension. Of these, clearly the appropriate description to which to reduce "purple swan" is [PURPLE AND SWAN]. But if we require for reduction extensional definition alone, rather than incorporating the reductions of other biological terms into our story, then any of these physical descriptions will be an equally appropriate candidate for reduction.

What is important to notice here is that the mode of terminological reduction endorsed by Hempel, extensional definition, is not exclusively susceptible to this argument. Any mode of terminological reduction that treats reduction on a term by term basis will be equally vulnerable. By ignoring the relation between the reduction of "purple swan" and that of its component terms, we leave something out about what reduction requires.

It might, of course, be the case that phrases like "purple swan," which are the logical combinations of discrete semantic parts, should not themselves be treated as biological terms suitable for reduction. That is, the above illustration may be problematic because of its treatment of phrases such as "primate with no tail" as biological terms in the same sense that "primate" and "tail" are biological terms. But one can always coin a biological term by reference to other biological terms, and define a new word, "pur-swan," as "any swan that is purple." The above considerations regarding the reduction of "purple swan" still ought to apply to the term "pur-swan"—it ought to be reduced to [PURPLE AND SWAN], rather than to some other physical property with which it just happens to be co-extensional. But although "pur-swan" is one syntactical unit—one word—it may still be argued that this word should be treated as if it were a combination of multiple semantic parts. Just as a physical reduction of the sociological term "bachelor" ought to be the logical combination of the physical reductions of "unmarried" and "man," so too for biological terms like "pur-swan." It may be claimed, therefore, that if a biological word is analytically equivalent to the combination of other biological terms, it ought not be treated as an atomic biological term that should itself be reduced.

Without a distinction between analytic and synthetic truth, however, such a position would be untenable. If all biological terms are partially defined by their relations to other biological terms, then, to some extent, all biological terms are similar to "pur-swan." A successful reduction of biology to physics is one in which, in addition to establishing extensional equivalences between biological and physical terms, also establishes the associations between closely related biological terms as physical necessities. For example, the physical reductions of "testosterone," "male," "human," and "growth" ought all be appropriately related so that, as a matter of physical necessity, TESTOSTERONE causes GROWTH in something that is MALE AND HUMAN. This sort of relationship is less important in the case of highly revisable biological sentences, such as "No swans are purple"—these may turn out to be physical contingencies. But, as discussed above, a true reduction of "turkey" ought to capture all (or at least most) of the important biological features of turkeys. This means that the physical description to which "turkey" is reducible ought to imply, by the laws of physics, the physical properties to which a turkey's important biological features are reducible.

The Reduction of Laws

The lessons we have learned from our consideration of Hempel's account of the reduction of terms may be applied to his account of the reduction of laws. Hempel claims that the reduction of biological laws to physical laws is a matter of the logical deducibility of the former from the latter. This deducibility cannot, as he notes, be established without the help of bridge principles that connect some biological terms or states of affairs with physical ones.[14]

Consider an ideal case in which this sort of reduction works precisely as Hempel intends, beginning with an example of a non-reductive explanation: Maggie walks into a room, arranges a pile of crumpled newspaper in the center, and lights it on fire. The temperature of the room then increases. Why does it increase? According to Hempel's account of scientific explanation, we can explain why this is the case by deducing this state of affairs from natural laws. One can explain the increase in temperature by means of a folk theoretic (or "theoretic") covering law of the form: If a fire is started in a room, and the fire is not extinguished, then the temperature of the room will increase. Given this law, and the initial conditions, 1) a fire was started in a room, and 2) the fire was not extinguished, we may logically deduce that the temperature of the room will increase. Therefore, by Hempel's account, we can explain the increase of temperature in the room by means of the fact that an unextinguished fire was started in the room, and the covering law, "when a fire is started in a room and is not extinguished, the temperature of the room will increase."

In order to satisfy M2, it must be the case that all biological facts (including the laws and generalizations of biology) are explainable by way of physical laws and principles. Assuming Hempel provides an accurate account of the above case, if certain laws and principles of physics implied the covering law given above, then we could use these physical laws to explain all of the cases in which that covering law could be employed in an explanation. Further, as Hempel equates the deducibility of a fact from laws with the explanation of that fact by means of those laws, a deduction of the covering laws would also explain the covering law itself in terms of physics.

Clearly, the folk law that a fire increases the temperature of the surrounding air could never have been deduced *a priori* from physics and chemistry. The help of bridge principles, which link concepts like "fire" and "temperature" to physico-chemical concepts, are required for such a deduction to take place. In this case, the relevant bridge principles would be something like "all cases of fire are cases of the physico-chemical reaction COMBUSTION" (where "COMBUSTION" stands for some physico-chemical characterization), and "all cases of an increase in mean molecular kinetic energy are cases of an increase in temperature." If the physical process of COMBUSTION implies, by the laws of physics and chemistry, an increase in the mean molecular kinetic energy of the surrounding area, then we have successfully provided a reductive explanation for the fact that fire increases temperature.

A crucial weakness with Hempel's account of a reduction of laws lies in his characterization of bridge principles. As he sees it, bridge principles often, and perhaps always, take the form of the generalization, "all instances of X are instances of Y, where X and Y are terms of the theories between which the reduction is taking place."[15] Extensional definitions are just a special case of bridge principles in which we can say both that "all instances of X are instances of Y" and that "all instances of Y are instances of X."

Applying what we have gathered from our consideration of extensional definition here, we can see how the problems of Hempel's account of terminological reduction extend to his account of the reduction of laws as well. Considering the physical property TURKEY, which we have said is co-extensional with the biological term "turkey," we can see that one can, by Hempel's view, construct the bridge principle, "all instances of 'x is a turkey' are instances of 'x is TURKEY.'" By Newton's Second Law, if an object has a mass of m, then a force of 5 N will accelerate it at a rate of (5 N)/m. By the laws of physics, then, all instances of TURKEY are instances of the physical property of NOBUFFALO, which is defined as

> x is NOBUFFALO iff x is ORGANISM AND [x would accelerate at a rate of (5 N)/m_1 if a force of 5 N were applied to it at time t_1 OR x would accelerate at a rate of (5 N)/m_2 if a force of 5 N were applied to it at time t_2 OR . . . x would accelerate at a rate of (5N)/m_n if a force of 5N were applied to it at time t_n]

As the following generalization is evidently true, Hempel's account will in addition treat it as a legitimate bridge principle: "All instances of 'x is NOBUFFALO' are instances of 'x does not eat buffalo for breakfast.'" We have just provided the bridge principles, "All instances of 'x is a turkey' are instances of 'x is TURKEY'" and "All instances of 'x is NOBUFFALO' are instances of 'x does not eat buffalo for breakfast.'" We have also shown that, by the laws of physics, something is TURKEY implies that it is NOBUFFALO. By Hempel's account, then, we have just given a reductive explanation of why turkeys do not eat buffalo for breakfast.

This counter-example to Hempel's account of a reduction of laws is based on the illegitimacy of the bridge principles involved. Yet there are additional problems that are not limited to the kind which are drawn from his account of terminological reduction. Using only intuitively legitimate bridge principles and physical laws, I provide below an illustration of the insufficiency of the deduction of laws for the reduction of laws. Although this example uses non-biological (as well as biological) terms, it may be regarded as an illustration of how a genuine counter-example to Hempel's view might proceed.

Maggie enters a room, as before, and begins to gather a pile of crumpled newspaper. She then drops to the floor, unconscious. Why is there no fire in the room? The air in the room is not nutritious,[16] and humans can only remain conscious in the absence of nutritious air for a minute or two. Further, preparing and starting a fire takes more than a minute or two (and cannot be performed while unconscious). From these folk/biological laws, we can deduce the covering law, "if the air in a room is not nutritious, then a human cannot start a fire in that room."

We can deduce this covering law from the laws of physics and chemistry, with the help of legitimate bridge laws, in a way that, intuitively, is not a reduction of our covering law. Above, I gave as an example of a bridge law, the principle, "All cases of fire are cases of COMBUSTION." In addition, we shall also use the bridge law, "All cases of a room with non-nutritious air—that is, air which lacks the positive characteristics necessary to sustain life—are cases of an enclosed area in which the air contains no oxygen" (see endnote 16). The physico-chemical process of COMBUSTION requires a supply of oxygen. And so, by the laws of physics and chemistry alone, all cases of an enclosed area in which the air contains no oxygen are cases in which COMBUSTION cannot occur. Given this, one can deduce the covering folk/biological law that a human cannot start a fire in a room in which the air is not nutritious.

That the preceding paragraphs do not provide a reduction of this covering law is just as apparent as the fact that we could not explain why Maggie failed in her attempt to start a fire by the fact that COMBUSTION cannot occur without oxygen in the air. Laws are, by Hempel's view, the devices of a theory that are invoked when giving a theoretical explanation of some state of affairs. The fact that covering laws about what humans can and cannot do without oxygen explain Maggie's failure in starting a fire, while the bridge principles and physical law discussed above do not explain Maggie's failure, shows that the latter are not an adequate reduction of the former.

While the fact that COMBUSTION cannot occur without oxygen does tell us that Maggie will fail to start a fire, it does not tell us *why* she, in fact, does fail. The folk/biological law that successfully explains Maggie's failure is derived from a number of other folk/biological laws, namely: "A human will fall unconscious in a minute or two in absence of nutritious air" and "A human must be conscious for more than a minute or two in order to prepare and start a fire." To successfully reduce our covering law, these more fundamental laws must have their own reductions incorporated. In other words, it may be possible to reduce the covering law in question to physics and chemistry, but this will have to be done in a way which makes use of the fact that a HUMAN (given some physico-chemical characterization), by the laws of physics, must have a breathable supply of oxygen available to remain CONSCIOUS long enough to perform the task at hand.

This seems closely related to the conclusion drawn regarding terminological reduction—that reduction cannot be done on a term by term basis. Here, we see that, when reducing a law, it essential that this reduction be carried out in a manner that incorporates the more basic laws from which the covering law is derived. This seems to entail a conclusion for a reduction of laws similar to the conclusion of the preceding section regarding a reduction of terms.

It may be the case, however, that there are genuinely basic laws of biology, which are not derived from any others, but from which all others may be derived. Were this the case, their reduction could be treated individually and then the reductions of all derived laws would trivially follow. This is related to the objection voiced above to treating words like "bl-swan" as terms in their own right, suitable for reduction in the same way as "black" and "swan" may be. This objection, it was said, relies on the controversial analytic/synthetic distinction. In the case of laws, however, nobody would want to say that laws cannot be logically derivable from one another, so it seems as though this picture of reduction could work. It might be that a reduction of laws could successfully proceed one law at a time, just so long as we are careful to distinguish between the genuine basic laws of biology and mere derivatives of these laws which themselves are not candidates for reduction.

This is a question that I will here leave somewhat open-ended, although I suspect that a reduction of laws, like a reduction of terms, cannot be carried out one law at a time. While it is clear that the law that a human cannot light a fire in a room with non-nutritious air can be derived from other folk/biological laws, it is unclear why these others ought to be treated as more basic. While, in fact, this covering law was obtained by derivation from other laws, there seems to be no reason to believe that this could not have happened the other way around. One might, for example, observe the inability of Maggie and other humans to start a fire in a room with non-nutritious air and, knowing that humans can remain conscious without air for a minute or two, conclude that preparing and starting a fire takes more than a minute or two. This appears to indicate that, as a reduction of the laws of biology to physics needs to preserve the logical relationships between biological laws, one cannot carry out reduction one law at a time, but rather must consider the reductions of the laws to which it is logically related.

Notes

[1] Carl Hempel, "Theoretical Reduction," *Philosophies of Science: From Foundations to Contemporary Issues*, ed. Jennifer McErlean (Belmont, CA: Wadsworth, 2000).

[2] Hempel 470.

[3] Hempel 470.

[4] Hempel 470.

[5] Hempel 470.

[6] Hempel 470.

[7] W. V. O. Quine, *Word and Object* (Cambridge: Cambridge UP, 1960) 271-72.

[8] "Law" is here intended in its broadest and most neutral sense. As laws have traditionally been viewed as the explanatory devices employed by theories to explain the phenomena that they do, I have adopted use of this term here. However, R2 may be read in such a way as to call for the reduction of whatever theoretical devices are employed in an explanation provided by that theory. In this sense, Ronald Giere's *principles* would be just as suitable a candidate as natural laws for satisfaction of R2. See Ronald Giere, "The Skeptical Perspective: Science Without the Laws of Nature," *Philosophies of Science: From Foundations to Contemporary Issues,* ed. Jennifer McErlean (Belmont, CA: Wadsworth, 2000).

[9] Hempel 471.

[10] Hempel 471.

[11] Hempel 471.

[12] Hempel 471-72.

[13] t_1 is here intended to specify an instant. If one has metaphysical discomforts with instants, one may instead consider t_1 to be of very small but non-zero duration. By this reading, $t_1$must be precise enough so that turkey's mass does not fluctuate by more than one half of a nanogram in the interval to which t_1 applies. For instance, it would be illegitimate to claim that a turkey had some particular mass, specified to the nearest ten nanograms, on May 29, 2001, as its mass will fluctuate significantly during this time.

[14] Hempel 472.

[15] Hempel 472.

[16] "Nutritious" should be taken as equivalent to a biological characterization such as "possessing the positive characteristics required of air to sustain homeostasis in humans." It is imperative for this example that "nutritious" be defined in terms of the possession of *positive* characteristics. For instance, just as a glass of milk laced with poison will still posses the calories and vitamins necessary to be *nutritious* (as opposed to being *edible*), so too may some oxygen-rich air with a toxic chemical in it be considered *nutritious* (as opposed to being *breathable*). Nutritious-ness is a perfectly legitimate

folk biological property, which can be characterized without any appeal to the chemical composition of air. Adding poison to food, as is evident without any appeal to caloric or vitamin content, does not eliminate its nutritional value. If it merely did this, poison would be a diet aid, rather than a means of killing someone. Similarly, taking a whiff of non-nutritious air will merely leave one short of breath. Breathing toxic air will kill you.

Works Cited

Giere, R. "The Skeptical Perspective: Science Without the Laws of Nature." *Philosophies of Science: From Foundations to Contemporary Issues.* Ed. Jennifer McErlean. Belmont, CA: Wadsworth, 2000. 180-189.

Hempel, C. "Theoretical Reduction." *Philosophies of Science: From Foundations to Contemporary Issues.* Ed. Jennifer McErlean. Belmont, CA: Wadsworth, 2000. 470-476.

Quine, W. V. O. *Word and Object.* Cambridge: Cambridge UP, 1960.

The Precritical Kant and So Much More

Critical Commentary on Martin Schönfeld's
The Philosophy of the Young Kant: The Precritical Project (Oxford: Oxford UP, 2000)

Jennifer K. Uleman, *University of Miami*

This is a truly wonderful book.

I confess I hesitated to agree to adding to my workload reading and commenting on Schönfeld's *The Philosophy of the Young Kant.* And I suppose I wasn't sure what I would get out of reading about Kant's precritical writings. Truth be told, I am still not sure that *I* want to read all of Kant's precritical works themselves, but I am very happy that Schönfeld did, and I am very happy to have read Schönfeld's book, which I recommend heartily to all of you, whether or not you work on Kant.

"Whether or not I work on Kant?" That is a bit much, isn't it? In fact, no. Schönfeld's own introductory descriptions of what the book sets out to accomplish include setting the record straight on Kant's intellectual development, bringing attention to Kant's considerable precritical philosophical and natural scientific achievements, pointing up illuminating continuities between Kant's precritical and critical works, and motivating the crisis that led Kant to critique. These descriptions of that book's aims fail to mention that his book will also bring the reader up to speed on the entire intellectual climate in which Kant found himself. The book does so by discussing Kant's engagements with and contributions to that climate—a climate in which there was a lot going on. For example, Cartesians and Leibnizians debated whether there were two kinds of matter, living and dead, the forces and mechanics of which had to be described by correspondingly different principles—a debate Kant entered with his first published paper, written when he was twenty-three years old, "Thoughts on the True Estimation of Living Forces." The "metaphysicians" and the "mathematicians" debated divergent approaches to nature—one group committed to irreducibly qualitative differences among parts of nature, the other to nature's uniform quantifiability, one persuaded that mathematical descriptions of nature were doomed to remain "artificial" and partial, the other persuaded that mathematics is the descriptive language for things in themselves. Debates arose about kinds of causality–-mechanistic, teleological, and so on—and the proper ways to investigate each. This period also witnessed arguments about the precise role of God in the world: creator, yes, and sustainer too. But how? Did God wind the watch? Did He tinker with it? Did He patch things up after messy miracles—miracles designed, after all, by Him to let us know about His existence? Moreover, questions were raised about the ultimate purpose, or *telos*, of nature: Is nature's *telos* to reveal God? To sustain human existence? Or is nature's end simply the joint order and diversity of nature itself?

There was more: worries that a denial of mind/body interaction was tantamount to denying sin, which of course depends on sinful animation of the flesh: philosophers' dismay at Newton's lack of engagement with metaphysics, evidenced by a shruggy willingness to invoke God as needed; a proliferation of "physico-theologic" treatises arguing for God's existence from the designs of, among other things, rocks, thunder, fire, water, snow, grass, and bees. There were discussions of the isomorphism between logic and ontology. There was racism, Kant's own and that of the European Enlightenment in general, which Schönfeld discusses unflinchingly. Whether discussing central debates or reporting on local skirmishes, Schönfeld always tells enough that one can understand what is at stake and, for those readers not familiar enough with Kant's (or Leibniz's, or Newton's, or Wolff's) work to guess, Schönfeld elegantly describes ramifications. *Anyone* interested in early modern philosophy, or in any field that owes the terms of its problematics to early modern philosophy, to say nothing of anyone interested in Kant—pretty much anyone that is—should buy this book.

Before I conclude the paid portion of my remarks [smile], let me also mention how lively, and how full of truly engaging detail the book is. I learned not only about 18th century physics and metaphysics, but also about the Lisbon earthquake of 1755, about the only known female German philosopher of the age, Johanna Charlotte Unzer, about tides and coastal winds and the slowing of the earth's rotation, and more.

We have been asked to pose questions for the author. I have four. Two are rather technical, and ask for pointers on understanding critical developments in light of precritical claims, and two are quite general, asking about philosophical projects, overall.

1. The first technical question has to do with Kant's willingness to regard teleological causation as unproblematic. Schönfeld writes,

> [Kant] assumed the divine imposition of goals occurred in terms of final processes immanent to nature instead of external divine interferences.[1]
>
> [Kant] identif[ied] the causal vehicle of purposive events with the efficient causation of physical processes.[2]
>
> [F]or Kant . . . matter actually contained an urge to organize itself.[3]

As Schönfeld describes it, this urge was meant to work itself out in terms of attraction and repulsion and was describable by the laws of nature. So for Kant, teleological self-organization, far from disrupting or competing with mechanism, was written into the script of nature itself. Here is my question: The wills of all living things, including the wills of animals and other non-rational creatures, cause the realization of objects through representations of those objects. They do this because the representations, as goals or ends, guide action. How might this seemingly teleological causation of the will fit into nature?

2. The second technical question has to do with Kant's ultimate resolution to the problem of determinism and freedom. Can consideration of Kant's early work point us toward the preferability, for Kant, of either a two-world or two-aspect solution to the problem of freedom and causality? Or does it point to neither of these?

3. The third question has to do with the conclusions Schönfeld draws from his study. At the end, we see Kant writing his review of Swendenborgianism, *Dreams of a Spirit-Seer*, which Schönfeld, I think aptly, reads as Kant's own half-laughing, half-crying *reductio* of his own precritical dream of integrating the material and spiritual worlds into a single ontological reality. We see Kant abandon the dream of grand synthesis that characterizes the precritical period. Should we read this abandonment as a failure? And is the critical philosophy itself, complete with transcendental idealism and distinct phenomenal and noumenal realms, also a failure–-a brilliant one, to be sure, but a failure? Or is it truly a move into bigger and better things? To put the question another way, should we regard Kant's critical philosophy as a failure, if one entirely inevitable or at least well motivated by the philosophical problems facing Kant? Or do the motivations to transcendental idealism still apply today?

4. The fourth question is the most general. There was, for me, something unsettling about reading about the 23-year-old Kant, trying, if unsuccessfully, to broker a peace between competing and seemingly incompatible views. There was something unsettling in reading about his early advocacy of "Bilfinger's rule," namely, the rule that

> . . . if men of good sense, who either do not deserve the suspicion of ulterior motives at all, or who deserve it equally, maintain diametrically opposed opinions, then it accords with the logic of probability to focus one's attention especially on a certain intermediate claim that agrees to an extent with both parties.[4]

What was unsettling was that these facts about Kant's intellectual biography threatened to subordinate Kant's arguments, including, ultimately, his critical arguments for transcendental idealism, to his own "peace-maker" tendencies. I thought: perhaps this threat, the threat that life will be drained out of the arguments themselves, diverted into biography, or psychology, or historical contingency, is why so many philosophers resist the history of philosophy. I wonder what Schönfeld thinks about this, and about the benefits and dangers of doing history of philosophy in general.

I think that is all. I am thankful to Schönfeld for writing an invaluable book.

Notes

[1] Martin Schönfeld, *The Philosophy of the Young Kant: The Precritical Project* (Oxford: Oxford UP, 2000) 107.

[2] Schönfeld 107.

[3] Schönfeld 111.

[4] Quoted in Schönfeld 59.

Works Cited

Schönfeld, Martin. *The Philosophy of the Young Kant: The Precritical Project.* Oxford: Oxford UP, 2000.

Dreams and Freedom

Critical Commentary on Martin Schönfeld's
The Philosophy of the Young Kant: The Precritical Project (Oxford: Oxford UP, 2000)

Byron Williston, *Wilfrid Laurier University*

Schönfeld's book on the young Kant has an argument I find unassailable, namely that reality is, for the precritical Kant, coherent and unified. This is of course at odds with the fundamental assumption of the critical philosophy that there is a basic breach in reality between the noumenal and phenomenal realms. So, rather than try to upset this elegant way of presenting the distinction between early and late Kant, I want instead to look at two more particular interpretations of the Young Kant. The first concerns the nature of what Schönfeld takes to be Kant's *self*-critique in the *Dreams of a Spirit Seer*; the second concerns the problem of freedom, and especially its connection to morality.

Dreams of a Spirit Seer

Kant wrote *Dreams* in 1765. One of the reasons I want to look at this book is because it is so bizarre, so difficult to interpret, and, for anyone who has laboured over the prose of the first *Critique*, such a delight to read. The book was ostensibly an attack on the weird angelology of Emmanuel Swedenborg. In the preamble to this book, Kant predicts that the reader will be completely satisfied with what he has to say: "[f]or the bulk of it he [the reader] will not understand, parts of it he will not believe, and as for the rest—he will dismiss it with scornful laughter." Swedenborg, evidently, believed himself capable of conversing with the angels and the dead. Furthermore, the dead themselves are said to form a society of spirits organized into the form of a Great Man. That is, each spirit occupies a place, equivalent to a bodily organ, within a larger spirit-body. These spirit-bodies then occupy the place of yet larger organs, this time belonging to the Greatest Man. Significantly, the whole show takes place in the context of a world that looks just like ours, complete with gardens, galleries, and arcades.

For Schönfeld, it is this last feature of Swedenborg's vision that is troublesome. He writes, " . . . Swedenborg's world of angels is the ultimate and absurd consequence of Kant's own precritical project."[1] So, Kant's critique of Swedenborg is ultimately a self-critique. Why? Schönfeld's master-argument, as I have already hinted at, is that Kant's precritical project is defined by the attempt to wed two seemingly incompatible philosophical vantage points: Newtonian mechanics on the one hand; purpose, human freedom, the immortality of the soul, and the existence of God, on the other. Here, the immortality of the soul is most relevant. But Kant's precritical take on the soul is deeply ambiguous. That is, Kant seems committed to the claim both that the soul is somehow of a material nature—as distinguished from *being* matter—and that it is immortal. Here, then, is how Schönfeld expresses Kant's dilemma:

> [T]he inevitable consequence of the precritical project was that bodies and souls, or material and immaterial substances, are subject to the same laws. At the same time, the precritical project must not rule out the possibility of an after-life. . . . Because souls are substances that obey the same fundamental laws as bodies, the immaterial community of the souls must contain the same structure as the physical world. The *reductio ad absurdum* of the precritical project is Swedenborg's spirit-world—-a world whose inhabitants are not even aware of their postmortal state because it looks and feels just like their old home.[2]

This is a compelling interpretation, but do we need to go this far? Let me suggest a way in which Kant might have resisted this conclusion from within the confines of the precritical project. Kant makes a distinction between two ways of conceiving of the soul. He agrees with those who argue that the soul is not matter, but insists that the soul is nevertheless of a material nature. As Schönfeld points out, being of a material nature means that the soul must be an "elementary wellspring of force," and even that souls must be "subject to the same fundamental patterns of reality deduced in the *New Elucidation*."[3] From this, it is supposed to follow that Kant cannot distinguish his account of spirit-life from that of Swedenborg. But *does* this follow? The critique of Swedenborg is obviously not aimed at the general premise that there exists an afterlife. With this Kant agrees right until the end of his career. The attack, rather, is aimed at Swedenborg's claim that he has *experience* of the afterlife, very detailed experience. Kant clearly thinks that this is impossible, and ridicules it accordingly

In other words, I don't see any reason to suppose that Kant has, because of his claim that we must understand soul-substance on analogy with body-substance, painted himself into the corner Schönfeld puts him in. He does not need to say that the postmortal state looks and feels just like this one. The description of souls as elementary wellsprings of force is just too general or sparse to warrant that claim. Perhaps an analogy would clarify my point. Imagine Mary, born deaf. It might be that in the course of her interaction with other people, she came across frequent reports of strange things called sounds, for example the sound of trumpets. Suppose she has good reason to think that those who speak of "the sound of trumpets" are generally reliable and not given to deceiving her. She might then conclude with good reason that something called "the sound of trumpets" exists. She might even explain this to herself in the form of a transcendental argument: the condition of the possibility of my friends being non-deceitful (or, more basically, being my friends) is that such reports generally refer to existent things.

Curious about sounds, she asks her friends what they are. Told that the sound of trumpets is like seeing bright red, she remains unenlightened because this does not sufficiently *distinguish* sounds from sights. Unperturbed, her friends go technical: they tell her that the human ear is an astonishing transducer that transforms the energy of a sound wave into a compressional wave in the inner ear. The energy of this wave is then transferred into nerve impulses that in turn can be transmitted to the brain. Finally—this is the tricky part—the same psychophysical laws that govern the production of other *qualia* result in the hearing of a sound. Mary is now slightly more informed about the physiology of the ear and brain and knows that the explanation of sound is analogous to those of the other senses. But, clearly, she is still utterly perplexed about the phenomenology of sounds, and nothing in any of the explanations she has received will allow her to speculate accurately about them.

If this is right, then Mary's knowledge of the laws of nature governing hearing vastly underdetermines any knowledge she might have of the experience of hearing. By analogy, then, if all our knowledge of the soul is *as embodied*, then when it becomes disembodied—and even if we know that it is an elementary wellspring of force—we will not have a clue what the after-life will be like. This description of the soul also vastly underdetermines our knowledge of post-mortal existence. This is all Kant needs to say in order to distance himself from Swedenborg's hallucinations.

Freedom and Morality

Next, I want to look at the problem of freedom and its connection to practical philosophy. Schönfeld has a very clear and penetrating chapter on the *New Elucidation*, that text where Kant tries to solve the problem of free will. The driving theme here, as throughout Schönfeld's book, is to show that Kant is trying to reconcile metaphysics and science. More specifically, the problem is freedom and determinism. Kant's compatibilism is, in brief, as follows. There is a chain of events starting in the external world, continuing to motives, thence to will, thence to action. But Kant breaks this chain in half, claiming that the chain leading from the world to motives is determined efficiently, while the chain leading from motive to the will and ultimately to action is determined spontaneously. If this is right, then we have both spontaneity and efficient causation in a single chain, and freedom—as spontaneity—is not a problem. But of course it is problem, precisely because the relation between intellectual motives and the inclinations of the will has not been specified. If action is to be truly spontaneous, and therefore free, the will must be able to range over the possibilities presented to it by the motives. Kant must say this so as to avoid falling into the trap of determinism. As Schönfeld points out, whereas the motive is passive, Kant sees the will as active, i.e. it ultimately causes itself. Freedom consists in self-determination.

The result of all this is that Kant preserves both rational freedom and necessity. Schönfeld's question, however, is this: does Kant succeed in deriving both from a common ground?[4] It is important to see why the answer to this question is "no." We might say that a common ground of both kinds of events is the principle of sufficient or determining reason. Clearly, the chain of efficient causes leading to the formation of motives obeys this principle, but so too, it might be argued, does spontaneous action since it is grounded in the self. But Schönfeld very astutely notes that this option is not available to Kant because he can give no content to the notion of self-determination. My free choices are *not* the product of reason, as the later Kant will say, because rational motives belong entirely to the chain of efficient causes. There is no common ground between efficient and spontaneous causation because there is no determinate ground at *all* to spontaneous causation. I want to relate *this* problem to Kant's early ethics.

Schönfeld points out that in the *Critique of Pure Reason*, Kant denounced the project of the *New Elucidation* as an impossibility.[5] In the later Kant, freedom becomes a practical postulate. This leaves one with the impression that Kant does not have too much more to say in the precritical period about the problem of human freedom. But already within the precritical corpus there is a pronounced move toward the critical position on this issue. Two texts in particular are interesting in this connection, namely, *Observations on the Feeling of the Beautiful and the Sublime* (1763) and the *Inquiry Concerning the Distinctness of the Principles of Natural Theology and Morality* (1763). Now Schönfeld does discuss these texts—though sparingly—observing, for example, that with the *Observations* " . . . metaphysics had been cast out, and practical philosophy had usurped its throne."[6] What I want to suggest, however, is that Kant actually delivers in this text a foundation for human freedom that can serve as an alternative to the metaphysical foundation sought after in the *New Elucidations*.

The *Observations* and the *Inquiry* are Kant's attempt to come to terms not only with Rousseau but also with the British moralists, most notably Hutcheson. The *Observations* makes a distinction between the beautiful—feminine, joyous, agreeable—and the sublime–-masculine, earnest, committed to principles, and so on. As far as the emotions are concerned, sympathy and complaisance (*Gefälligkeit*) are beautiful, while the dutiful subordination to moral principles is sublime. For Kant, the moral feeling of the sublime is the feeling of the beauty and dignity of human nature. This feeling then comes to form the bedrock of moral obligation. It is the immediate effect of the consciousness of the feeling of pleasure combined with a representation of an object. However, as Alfred Denker has argued, this left Kant dissatisfied. He could not decide whether reason, through the formulation of a necessary end of action, constructs the contents of moral feelings or whether the notion of a necessary end is an unanalysable constituent of moral feeling. Denker states the choice this way: we have here either an ethics of "autonomous practical reason" or an "intuitive ethics of value."[7]

While Kant himself does not answer this question, it is important to note that whatever answer he might give to it would go a good way toward solving the major problem of the *New Elucidations*. Recall that this is the problem of finding a determining reason for free action. In the *Observations*, Kant claims that the object of moral feeling (the sublime) is different in kind from the object of moral sympathy (the beautiful). More specifically, the ground of moral feeling is universal and independent of our subjective inclinations. The fundamental ground of obligation is rooted in the sublime apprehension of the dignity of human nature. But this now provides a principle for spontaneous action as such, one that can both ground a practical conception of human freedom and remain sharply distinguished from the play of subjective inclinations. There are two reasons for thinking that this is what Kant was up to.

First, Kant's lecture course of 1764 on the philosophy of Rousseau was explicitly designed to instill in his students a passion for thinking for themselves. The following cluster of ideas recurs throughout the lecture notes. Philosophy is a therapy for the corrupted human condition. Philosophy teaches universal respect. The natural state of humans is to be free and equal. The implication of all this is clear: we are equal because we are all free, and the highest expression of our freedom is to treat others in accordance with their intrinsic dignity, i.e. equally. It is therefore our duty to oppose the injustices that are solely a product of artificial inequality.[8]

Second, and more decisively, in the *Inquiry* Kant marks off his moral philosophy from that of Crusius in at least one important respect, namely, in the distinction between *necessitas legalis* and the command of God. Crusius believed that morality is externally rooted in the will of God and that moral motivation is therefore predicated on obeisance to this external source. Kant, by contrast, argues that morality is a deliverance of human nature itself. This move encapsulates a trend in the early history of modern moral philosophy, away from divine command theory and toward a notion of autonomous self-governance. Previous natural law theories—those of Hobbes and Cumberland, for instance—held that force, in the form of sanctions and rewards, is a legitimate ground of obligation. Clearly, Kant is already rejecting this theory.

If these reflections are on the right track, then one may have to qualify the claim that, with respect at least to the problem of human freedom, the precritical project *as a whole* can be defined as the attempt to provide a *theoretical* reconciliation of science and metaphysics. Rather than attempting to find a theoretical ground of spontaneous action, Kant is already talking about the ground, located within practical judgment as such, of autonomy.

Again, these are minor points. Before reading Schönfeld's impeccably argued book I was one of those who thought that there was no single precritical project, but that the texts of that period were nothing more than a confused and confusing cocktail of philosophical opinions—some insightful, others crazy red herrings. Happily, Schönfeld has awakened me from *that* dogmatic slumber.

Notes

[1] Martin Schönfeld, *The Philosophy of the Young Kant: The Precritical Project* (Oxford: Oxford UP, 2000) 241.

[2] Schönfeld 244.

[3] Schönfeld 244.

[4] Schönfeld 159.

[5] Schönfeld 160.

[6] Schonfeld 231.

[7] Alfred Denker, "The Vocation of Being Human," *New Essays on the Precritical Kant*, ed. Tom Rockmore (Amherst, NY: Humanity Books, 2001) 139.

[8] Cf. Denker 147.

Works Cited

Denker, Alfred. "The Vocation of Being Human." *New Essays on the Precritical Kant.* Ed. Tom Rockmore. Amherst, NY: Humanity Books, 2001.

Schönfeld, Martin. *The Philosophy of the Young Kant: The Precritical Project.* Oxford: Oxford UP, 2000

Some Questions on Negation and Possibility

Critical Commentary on Martin Schönfeld's
The Philosophy of the Young Kant: The Precritical Project (Oxford: Oxford UP, 2000)

***Sidney Axinn**, University of South Florida; Temple University, Emeritus*

Professor Schönfeld had a difficult problem: how can you keep the attention of your readers when they already know the outcome of your story? His story is Kant's precritical project, and we all know that it failed, in Kant's opinion. (And we are glad that it failed—from the failure came the three *Critiques*, and the other significant work.) Amazingly, Schönfeld does keep his readers' attention. He keeps us interested even though we know that we are reading about empty metaphysics.

Schönfeld gives us a lot of philosophic material, not just anecdotes. He reconstructs Kant's argument in each paper or book that he considers, gives us the historical background and adds his own comments. I'll select just a few items here and there, and raise a question or two about them. But, nothing I say takes away from the fact that this is a very valuable study of Kant's precritical work, and should stand for a long time. Even those who are not persuaded by the thesis that Kant had a unifying project in the early work must be impressed by Schönfeld's analysis and supporting material and notes. Apart from the main theme of Kant's project, we learn a lot about Kant's intellectual environment. For example, Schönfeld gives us the only clear description I've ever seen of the five Bernoullis, the family of mathematicians and philosophers. Also, we find Christian Wolff's views in some detail. There are regular turns to Leibniz to find with what Kant did or did not agree. And many more very significant connections are made.

The Introduction gives generous space to Schönfeld's opponents, those who take the precritical period to be trivial, philosophically. He quotes Lewis White Beck's remark that prior to the critical philosophy "Kant . . . would deserve a quarter of a page in Ueberweg."[1] Actually, in the English language translation of Ueberweg (done in 1873), Kant gets 57 pages, of which there are eight pages devoted to the precritical period.[2] Since this is a very small print edition, eight pages are a lot of material. But, of course, Beck's point is that it is the critical works that give us any interest in the earlier writing. Actually, Ueberweg, himself, published on some of the precritical works, so this attention in his history is not surprising. Enough Ueberweg: back to Schönfeld.

Schönfeld surprises us by first mentioning Kant's "extraordinary honesty," and shortly after that telling us that Kant denied having a position in the Living Forces debate, although he really did have "a stake in the conflict."[3] Schönfeld also suggests that "perhaps in an effort to pretend greater originality, Kant emphatically rejected aspects of the standard Leibnizian position—only to argue for a view "almost indistinguishable" from it. It looks as if Kant is ordinarily human rather than extraordinarily honest. Of course, later on, Kant would tell us about the dear self that speaks in each of us.

Schönfeld's style is quite smooth. He remarks on the role that Newton gives to God, that of a supreme creator who later becomes a mere handyman, calling this a "bad career move for a god."[4] He also enjoys the pun on the author, named Seidel, who wrote on caterpillars secreting silk threads, which are s*eide* in German.

In the 1755 work, *New Elucidation of the First Principles of Metaphysical Cognition*, Kant tries to combine physics and freedom, as Schönfeld describes it. This involves, among other matters, the basic logical concepts, negation and possibility. I'll say something about each of these.

Negation

Kant held that, in Schönfeld's words, "True propositions . . . express both affirmative and negative contents."[5] Schönfeld follows this with an odd statement, "[b]ecause one can derive neither a negation from an affirmation nor an affirmation from a negation, the basis of all true propositions must express both."[6] It is not clear whether this is Kant's or Schönfeld's belief. In either case, it is odd because a statement, *A*, is equivalent to the denial of *not-A*. A denial counts as a negation. There is a certain ambiguity in the remark, "the basis of all true propositions must express both [affirmation and negation]." Does this mean that affirmation and negation say the same thing in different ways, or that they say different things? If they say different things, that is interesting and calls for a further analysis of the different contents.

At any rate, as Schönfeld says, this was a break with the tradition. It is an interesting one and, of course, varieties of negation continue to be seen. An article by this commentator,"[7] uses an idea similar, although not identical, to Kant's. In one version, a consequent of one interpretation of Kant's view is that a situation may be described by either an affirmative or a negative statement. (For example: The statement that I am not wearing a hat is equivalent to the statement that I am bareheaded.) Put another way, whether a statement is affirmative or negative depends on the context: no statement is either affirmative or negative independent of context. If this holds up, no properties are positive or negative, apart from a context. While this idea has not been generally adopted, one prominent logician, Quine, has spoken favorably about it. I don't think that Kant, himself, did much with it, from the logical standpoint; but the ontological proof for the existence of God hovers in the background of this position. If there are no independently positive or negative properties, the standard version of the Ontological Proof is in trouble.

In these comments about Kant's view of negation in this early work, I've drifted between considering propositions and statements. A careful discussion would have to be more consistent.

Possibility

Kant's view of possibility is particularly interesting. In *The New Elucidation*, Schönfeld quotes the following, somewhat mysterious, claim: "nothing can be conceived as possible unless whatever is real in every possible concept exists . . ." Then, in "The Only possible Argument in Support of a Demonstration of the Existence of God" (1763), Schönfeld tells us that Kant "hopes to show that the complete set of thinkable data is the complete set of all positive[?] properties that exists as a unified entity endowed with divine qualities." [8] Further on, we get Kant's premises, one of which is "The Material Condition of Possibility, Anything that is possible must be thinkable, and for anything to be thinkable, the presence of material data to the mind is required."[9]

Schönfeld takes the Material Condition to be "a fatal flaw." Why? Because "a conceptual analysis of 'possibility' reveals the possibility of a conceptual whole and the possibility of its conceptual elements—it does not reveal the possibility of a conceptual whole and an independent and prior existence of its conceptual elements. This is the fatal flaw."[10] Here Schönfeld is not as generous to Kant as he usually is. Schönfeld looks for the source of this material condition in Leibniz. That certainly is a reasonable place to go. But there is another move to consider.

In the first *Critique*, Kant has a few pages on possibility. In both the A and B editions he introduces the matter modestly.

> To enquire whether the field of possibility is larger than the field which contains all actuality, and the latter, again, larger than the sum of that which is necessary, is to raise *somewhat subtle questions* which demand a synthetic solution, and yet come under the jurisdiction of reason alone.[11]

Are there more possibles than actuals, or more actuals than possibles (or are they equal in number)? Common sense since Aristotle holds that there are more possibles than actuals. But Kant sneers at "the poverty of the customary inferences through which we throw open a great realm of possibility, of which all that is actual (the objects of experience) is only a small part."[12] Shortly after this he insists that "without material nothing whatsoever can be thought."[13]

Where does Kant go with this idea that there are not more possibles than actuals? Not very far. This section started with his remark that this topic raises "subtle questions." And he ends the section saying, "[w]e have therefore had to content ourselves with some merely critical remarks; the matter must otherwise be left in obscurity until we come to the proper occasion for its further treatment."[14] The analysis of possibility is so complicated that Kant didn't want to go into it in a book as simple and clear as the *Critique of Pure Reason.*

Although Kant dropped the topic, Nelson Goodman, the American philosopher of science, seems to have taken Kant's idea and developed it.[15] I'll introduce Goodman's view with an example. We have seen horses and wings, so we can imagine Pegasus, a winged horse. We can imagine the wings put on upside down, put both on one side, etc. But, were there something made of parts that we have never experienced, how could we tell it to anyone else, or even to ourselves? If we have never experienced something, we don't know its color, shape, size, material, texture, etc. Originality involves arranging and rearranging what we have or have experienced: originality can't hope to create from nothing. Even the greatest of art schools must have a supply store where the artists get their materials. (Even the creator, in Plato's *Timaeus*, had to start with something, the chaos, to make the world.)

In the vocabulary of the Critical Kant (and ours) one can name objects that cannot be thought. He distinguished between "real possibility" and "merely logical possibility."[16] Such names for objects that are not real possibilities, that are not to be thought of as in the phenomenal world, he famously takes to be in a noumenal world, if they are logically consistent. The names of noumenal entities can be mentioned grammatically, but not used to refer to anything in the actual world. Nor can they be thoughts for humans. Our ability to name is not an ability to create a possible object from nothing, from no parts.

Can there be a possible object that is not constructed of parts? If there were anything simple rather than compound, it could not be *constructed in thought.* If there are no parts to be analyzed, there are also no parts to be synthesized. Since possible objects are synthetic, an ultimately simple entity cannot be understood, on this basis.

For an example of this view of possibility, consider this. On one side of a road there are two auto engines, one red and one blue. On the other side of the road there are two auto bodies, one red and one blue. Assume that a complete auto requires an engine in a body. Now there are four possible autos, the red engine in either the red or the blue body, and the blue engine in either body. Any possible auto must be made of actual parts.[17]

This detour is meant to show that Kant's Material Condition is not a "fatal flaw," as Schönfeld puts it, but a powerful idea that Kant returned to in the first *Critique*, but did not develop in detail. The development was carried out by a 20th century philosopher, Nelson Goodman. Would the nominalist dichotomy of part and whole be congenial to Kant? That is a separate question.

Schönfeld has an objection to this version of possibility, if I understand the paragraph at the bottom of p. 203. He holds that "In this reading, Kant's thesis amounts to the assertion that existence precedes possibility." One might respond, "yes, it either precedes or is simultaneous with possibility." Where there are no parts, there are no possible arrangements of parts. Perhaps we can talk about that.

Conclusion

I conclude with the following questions and comments:

1) Did Kant have a single precritical project? He did write on other matters—the Lisbon earthquake, for example. But, after leaving out a few such miscellaneous items, Schönfeld does have a smooth story here.

2) If we lost the precritical work, would there be a serious loss? Some of the scientific writing has held up. The so-called "Kant-Laplace theory" would not be lost to us, although Kant's role would be. The philosophic constructions that are valuable are developed in the first *Critique*. So Lewis Beck is convincing: for his precritical work alone, Kant would deserve merely a paragraph in Ueberweg, at most.

3) What can we learn about Kant's mature work? I enjoyed reading the history of Kant's struggles with many things, for example, the several efforts to prove the existence of God. But, we can understand the analyses of these proofs in the first *Critique* without this history. And so for the other topics.

To stop here would be to miss the main point. This is a history of ideas, and in this way we can see how the mature philosopher developed. Also, the history of ideas is valuable for its own sake, and this is certainly a major contribution to that history.

Notes

[1] Martin Schönfeld, *The Philosophy of the Young Kant: The Precritical Project* (New York: Oxford UP, 2000) 6.

[2] Friedrich Ueberweg, *History of Philosophy: From Thales to the Present Time*, 4th ed., vol. 2, trans. Geo. S. Morris (New York: Charles Scribner's Sons, 1873).

[3] Schönfeld 19, 39.

[4] Schönfeld 105.

[5] Schönfeld 133.

[6] Schönfeld 133.

[7] Sidney Axinn, "Ayer on Negation," *Journal of Philosophy,* 65.2 (1964): 74-75.

[8] Schönfeld 195. This is incomplete since Schönfeld had given us Kant's view, above, that "the basis of all true propositions must express both [affirmation and negation]."

[9] Schönfeld 201.

[10] Schönfeld 205.

[11] Immanuel Kant, *Gessammelte Schriften,* ed. Akademie der Wissenschaften (Berlin: Reimer, 1900-1942) A 230/B283, emphasis mine.

[12] Kant A231/B281.

[13] Kant A232/B284.

[14] Kant A 232/ B 285.

[15] Nelson Goodman, "The Passing of the Possible," *Fact, Fiction, and Forecast* (Cambridge, Mass.: Harvard UP, 1954) 37-62. See also Sidney Axinn, "Kant and Goodman on Possible Individuals," *The Monist* 61.3 (1978): 374-385.

[16] Kant B xxvi.

[17] This is taken from Goodman's example, not Kant's.

Works Cited

Axinn, Sidney. "Ayer on Negation." *Journal of Philosophy* 65.2 (1964): 74-75.

Axinn, Sidney. "Kant and Goodman on Possible Individuals." *The Monist* 61.3 (1978): 374-385.

Goodman, Nelson. "The Passing of the Possible." *Fact, Fiction, and Forecast.* By Goodman. Cambridge: Harvard UP, 1954. 37-62.

Kant, Immanuel. *Gessammelte Schriften.* Ed. Akademie der Wissenschaften. Berlin: Reimer, 1900-1942.

Schönfeld, Martin. *The Philosophy of the Young Kant: The Precritical Project.* New York: Oxford UP, 2000.

Ueberweg, Friedrich. *History of Philosophy from Thales to the Present Time.* 4th Ed. Vol. 2. Trans. G. S. Morris. New York: Charles Scribner's Sons, 1873.

Response to Commentaries

On *The Philosophy of the Young Kant: The Precritical Project* (Oxford: Oxford UP, 2000)

***Martin Schönfeld**, University of South Florida*

I would like to thank Sidney Axinn, Jennifer Uleman, and Byron Williston for their insightful and generous comments. All three raise interesting questions that need to be answered. Before trying to do so, I should briefly explain what Kant's precritical philosophy involves and what *The Philosophy of the Young Kant* is about.

The Kantian *oeuvre* is divided into two periods. Famous is the period from 1781 to 1804, in which Kant wrote his most influential works, such as the three *Critiques*, the famous essay "What is Enlightenment?," the *Foundations of the Metaphysics of Morals*, and the prophetic treatise "On Universal Peace." His efforts prior to the critical turn, however, are largely unknown. I wanted to study these precritical writings because their obscurity puzzled me. Not even Kant scholars read them as a rule. This is strange, because Kant wrote *a lot* before the *Critique of Pure Reason* (1781)—a book on cosmology, the *Universal Natural History* (1755); a dissertation on first principles and free will, the *New Elucidation* (1755); a dissertation on elementary particles, the *Physical Monadology* (1756); a book on rational theology and metaphysics, *The Only Possible Argument in Support of a Demonstration of God's Existence* (1763); a treatise on aesthetics, *Observations on the Beautiful and the Sublime* (1764); a treatise on the methodology of philosophical research programs, *Inquiry concerning the Distinctness of the Principles of Natural Theology and Morals* (the so-called "Prize Essay" of 1764), and the obscure and tortured *Dreams of a Spirit-Seer* (1766). Up to the *Inaugural Dissertation* (1770), which initiated Kant's "silent decade," he had produced three books, half a dozen treatises, and (depending on how you count them) fifteen papers.

The neglect of the precritical writings is stranger still if one remembers that they were not half-cooked *juvenilia* but works of a thinker in his prime. One can arguably dismiss Kant's first book on *Living Forces*[1] as the flawed composition of a twenty-two year old, but to shrug off the works produced afterwards is not so easy. These were not products of immature youth. Kant wrote his second book (the *Universal Natural History*) in his early thirties and the third (the *Only Possible Argument*) when turning forty. These texts contain a number of startlingly accurate insights and discoveries. Consider merely their contributions to our current knowledge of nature. In the "Aging Earth" essay (1754), Kant figured out that the axial rotation of the Earth is slowing down until (far, far in the future) a terrestrial day will be as long as a lunar month. In the "Theory of Winds" (1756), he correctly identified the causes of the coastal winds, trade winds, the equatorial passat, and of the seasonal occurrence of the monsoon. In the *Universal Natural History*, he was the first to understand why planetary orbits are roughly arranged on the ecliptic plane, that the luminous smears visible on the night sky are galactic clusters, and how the present-day solar system originated from a cosmic cloud.

Why, then, this strange neglect? Mostly, it is Kant's own fault. The *Critique of Pure Reason* was for him a fresh start based on new insights, and he roundly rejected the works prior to the "great light" of 1769—going so far as resisting their reprint in the first collection of his works. Kant scholars never questioned the claim of their master that the early works were rubbish. I suspect that this is partly due to the comforting subtext of Kant's self-portrait: it *is* possible, the story suggests, to remain mediocre for most of one's life and hold off writing one's masterpieces until old age. Ernst Cassirer, in his 1918 study of Kant's life and works, articulated the standard assessment of the precritical period: the young Kant was an unoriginal scatterbrain who worked on obsolete problems and failed to develop a coherent view. In the *Philosophy of the Young Kant* I argue for the very opposite. I claim that the early Kant was an original and innovative thinker wrestling with timely issues and perennial questions, who systematically constructed an ambitious reconciliation of science and metaphysics. This construction of a "unified field theory," as it were, was Kant's precritical project. In my study, I give an account of the precritical project from its misguided beginnings in the 1740s to its development in the 1750s and to its culmination and collapse in the 1760s. I read the ontological dualism of the *Inaugural Dissertation* (1770) as the aftermath and result of the collapse, and accordingly end my inquiry with the text that acknowledges the failure of the precritical project, the *Dreams of a Spirit-Seer* (1766).

So much about the early *oeuvre* and my reading. Now to the questions, starting with the specific and technical, and ending with the general and philosophical.

Williston: Why did the encounter with the visionary Emmanuel Swedenborg doom the precritical project? Did Kant *have* to acknowledge defeat in the *Dreams of a Spirit-Seer*?

Professor Williston doubts that Kant has, because of his perceived analogy between body-substance and soul-substance, painted himself into a corner, and he is certainly right. An analogy between bodies and souls, or any compatibilist ontology, is not incoherent by default. What triggered Kant's recognition of defeat was not this analogy in particular, but rather a fatal mix of ontological and methodological concerns. Kant suggested with the *Universal Natural History* a model of physical reality based on Newtonian principles. In the *New Elucidation*, he proposed that the structure of physical reality is fundamentally compatible with the structure of the part of reality that is not physical—the "intelligible sphere" of free action, thoughts, souls, and God. In the *Only Possible Argument*, he tried to illustrate this unified ontology by constructing two parallel proofs of God, one about God's intelligible features, the other about God's mark on nature, and by showing how both derive from the same ontological presuppositions. In the *Prize Essay* he articulated the methodological constraints of a philosophy of nature based on unity and certainty.

Enter a Swedish seer who claims, in his wildly speculative *Arcana Coelestia* (8 vols., 1749-56), that the world is indeed ontologically unified and that he, Swedenborg, has access to both its physical and intelligible spheres. This self-styled visionary describes the "world of the angels," the intelligible sphere of the soul-substances, as a mirror-image of physical reality which, coincidentally, reveals the afterworld to look like a heavenly Stockholm. Kant recognizes an unintended caricature of his precritical project in the *Arcana Coelestia*, raising doubts about the verifiability of unified and compatibilist models in general. The crucial question, for Kant, is over knowledge. How can we substantiate any claims about the intelligible? In the *Prize Essay*, written before the encounter with Swedenborg, Kant proposes that the phenomenological certitude that accompanies the correct analysis of abstract concepts constitutes the decisive criterion of metaphysical truth. Then he reads the *Arcana.* He sees that Swedenborg trusts his visions; Swedenborg is evidently the proud proprietor of the inner certitude demanded from metaphysics—but he is nevertheless *wildly* wrong! In the subsequent *Dreams of a Spirit-Seer*, Kant recognizes that the *Arcana* highlights a flaw in his criterion of truth, undermining his hope for elucidating the intelligible and persuading him that the precritical project was overly ambitious.

> **Williston: Can the precritical project as a whole be defined as the attempt to provide a theoretical reconciliation of science and metaphysics? What about Kant's treatment of moral feeling in the *Observations on the Feeling of the Beautiful and Sublime* (1764)? Doesn't the notion of moral feeling solve the problem of a determining reason for free action in the *New Elucidation*?**

Professor Williston is perfectly right in reminding us that the early Kant did not only work on the theoretical reconciliation of science and metaphysics. The precritical project, as the quest for this reconciliation, is a subset of the precritical philosophy. But I claim that *most* of Kant's precritical ideas, questions, and texts, concern this reconciliation. I refer to twenty-one of the twenty-four writings up to and including the *Dreams*; that is, the entire *oeuvre* except the *Eulogy on Funk* (1760), the *Maladies of the Head* (1764), and the *Observations.* (In the pagination of the *Academy Edition*, I thus exclude 73 of 888 pages from 1747 to 1767 as irrelevant for the precritical project.) Kant's treatment of moral feeling as the universal ground of moral action reveals his interest in practical philosophy. But moral feeling does not help to solve the problem of freedom with which Kant wrestles. Ethically relevant actions presuppose responsibility and freedom from external constraints. Compatibilism is an attractive causal account because it acknowledges both the lawfulness of processes in the physical world and spontaneity in the ethical sphere. Yet, compatibilism seems to be impossible to demonstrate. An ontologically unified model of reality will marry free action to lawful process, and because the latter is deterministic, the freedom of the former dissolves. In the *Observations*, Kant flirts with moral feeling as the ground of free action. For action to be free, the ground must be a motive deliberately embraced by the will, not a compulsion subjecting the will. The appeal to moral feeling thus merely illustrates the precritical ontology without repairing its flaws.

Uleman: How might the teleological causation of the will of living beings, such as animals, fit into nature?

Professor Uleman observes that living beings cause the realization of objects through their representations, and that these representations guide action as ends. There is thus a teleological causation of the will. The conception of teleology prevailing in the early 18th century (shared by Wolff, School-Philosophers, Newton, Pietists, and Physico-Theologians) had stipulated that God imposes purposes on nature from a supernatural vantage point. Kant rejects this because he finds it incompatible with the causal structure of physical processes. Divine interferences would create effects in nature without natural causes; they would remain inexplicable miracles, violating the law of cause and effect. In their stead, he suggests that purposive developments must obey the natural patterns of causality. For Kant, teleological self-organization (to borrow Professor Uleman's phrase) is written into the script of nature itself.

The teleological causation of the will of living beings conflicts neither with the purposive developments of natural systems nor the lawful regularity of physical processes. According to Kant's immanent teleology, both kinds of final causation (of living beings and of natural systems) mesh with the efficient causation governing physical processes. A purposively realized event is genuinely caused, and its cause is within nature, either as the telic striving of natural systems or as the goal-directedness of animals. Kant also avoids the trap of retroactive causation. Final causes do not act backwards through time but precede their effect in both kinds of teleological process. The striving of natural systems antedates the event of unfolding that brings the purpose of nature's self-perfection about; likewise, the will of a living being leads, as intended goal, to the action that then furthers this goal.

The immanent teleology is the driving force behind the precritical project. As the glue bonding physical nature with living beings, it suggests to the young Kant a way of reconciling efficient and final causation and motivates him to attempt their reconciliation with spontaneous causation as well (although the latter venture failed). Still, Kant is a child of his time. He discusses organic nature only in passing, and his attention on inanimate nature reflects the fact that the Newtonian revolution of physics dominated his age. We now know that inanimate nature cannot be described in terms of purposes. Nonetheless, this projection of purposes on physics was heuristically useful, for by regarding nature as a directively organized system Kant came up with a number of actual astrophysical discoveries. Particularly his revision of final causation as an immanent teleology seems to have withstood the test of time. The failure of reductionist philosophy of science has reinforced the claim of various philosophers of biology about the methodological value of functionalist explanations in the life sciences. This failure has also led to a revival of immanent teleology in current environmental ethics as the best explanation of the phenomenon of life and the most compelling case for life's intrinsic value.

Uleman: Does a consideration of Kant's early work point toward the preferability, for Kant, of either a two-world or two-aspect solution to the problem of determinism and freedom? Or does it point toward neither of these?

In the *New Elucidation*, the early Kant fails to achieve a compatibilist resolution of the problem of determinism and freedom by means of a unified ontological theory of causation. In the *Inaugural Dissertation*, he cuts through the Gordian knot of the hoped-for unified ontology and slices the world into two halves, a deterministic and physical *mundus sensibilis*, and a free and conceptual *mundus intelligibilis*. The bifurcation of reality into the empirically accessible phenomenal realm and the inaccessible noumenal sphere remains his definitive position. Seen in this light, the fate of the early work points us directly to the positions Kant advocates later. The critical philosophy was Kant's attempt to make the best of his previous defeat. The two-world/aspect solution, in the critical period, is the result of his inability to solve the problem of determinism and freedom. The two-world view, then, is a concession of defeat. At best, it is a heuristic assumption, a stepping-stone to an eventual resolution of the puzzle, for judged as a genuine solution, the two-world view of the critical Kant is utterly unsatisfactory: the reality of deterministic processes and free actions is acknowledged, as well it should be, but we know little about either. Deterministic processes supposedly govern nature, but nature is merely the arena of our representations that are as empirically real as they are transcendentally ideal. Ethics presupposes freedom, but freedom is merely an unproven and unprovable regulative idea. We need to do better than this.

Axinn: Why do you consider Kant's "Material Condition" as a fatal flaw, particularly in light of Nelson Goodman's development of a very similar idea?

The "Material Condition of Possibility" is an assumption needed for the demonstration of God's existence in the *Only Possible Argument*; that is, the idea that anything that is possible must be thinkable, and for anything to be thinkable, the presence of material data to the mind is required. Kant infers from this that something must exist if anything is possible, and given that possibility cannot be negated, it follows that it is impossible that nothing exists. This, in turn, suggests to him that something must exist no matter what, and that, therefore, something exists necessarily. Now Kant has all he needs to complete his demonstration and to show that an *ens necessarium*, a divine necessary being, exists. Why can we not pull the divine rabbit out of a conceptual top hat? In *The Philosophy of the Young Kant*, I identify the location of Kant's error in his interpretation of the Material Condition of Possibility. Possibility is instantiated in possible concepts, and possible concepts are determined by predicates. Nothing can show us that such predicates have to exist *prior* to the possible concept itself. They have to, however, for Kant's argument to work. Accordingly, I argue that therein consists the first fatal flaw of the argument.[2]

Professor Axinn refers to Nelson Goodman's doctrine that the only possible entities are actual ones, a quasi-empiricist doctrine evidently resembling the Material Condition.[3] In his "Kant and Goodman on Possible Individuals,"[4] Professor Axinn argues that Kant's critical view on possibility, articulated in the "Postulates of Empirical Thought" of the *Critique of Pure Reason*[5] dovetails with Goodman's doctrine. Axinn explains there that this doctrine is more plausible than it might seem, for both Goodman and the critical Kant are concerned with constructions relative to human systems of understanding, and both authors would agree that a possible individual is composed of actual, experienced parts.[6]

I think Professor Axinn is right in chiding me because I dismissed the Material Condition too quickly. Admittedly, the Material Condition is not flawed if taken as a claim similar to Goodman's doctrine, as a logical condition relative to human language systems. Interpreted epistemically and in relation to human constructions, the Material Condition makes sense. But I am afraid this does not let the precritical Kant off the hook. Kant employs the Material Condition with ontological intent, using it, in a Leibnizian way, as referring to a *regio idearum* that posits experienced data as Platonic predicates. This changes the reference of the condition from relative to absolute possibility, and I doubt that Kant can get away with it.

Axinn: If we lost the precritical work, would it be a serious loss?

The precritical work compares to the *Critiques* like Karl Marx's *Paris Manuscripts* relate to the later *Capital.* Just as the early *Paris Manuscripts* shed light on the motivations and rationales of the mature Marx's masterpiece, the precritical *oeuvre* helps us to better understand Kant's critical philosophy. The critical Kant is notorious for his architectonic proclivity, which appears as a puzzling Prussian obsession with tidiness and order. Viewed in the context of the precritical *oeuvre*, however, this seeming obsession turns out to be a legitimate longing for a system of knowledge that integrates data in logical fashion. The critical system was supposed to deliver what the precritical project promised. Acquaintance with the precritical project also instructs us about the cornerstones of the critical system and their underlying rationales. Why did the critical Kant embrace dualism? Because he had learned the hard way that monism has intractable problems. Why did he relegate the ideas of God, soul, and world to the transcendental dialectic? Because he was familiar with the shortcomings of their constitutive employment through personal experience. Why did he subscribe to such an optimistic view of humankind's evolution? Because this envisioned progress is a corollary of the precritical teleology.

Like the *Paris Manuscripts*, the precritical *oeuvre* also has merits of its own. Consider the already mentioned scientific discoveries. Because the early works remained virtually unknown, each discovery was effectively made twice, first by Kant, then by others in later times. Now we "have" these discoveries, and if the precritical *oeuvre* were lost, nothing would change. But this does not undermine the significance of Kant's insights. Johannes Gutenberg's invention of the printing press in 1434 has "given" us the printed word. This does not lessen the importance of earlier inventions of movable type by the Chinese and Koreans. Bentham and Mill "gave" Western ethicists the greatest happiness principle. But this does not reduce the value of the same utilitarian idea conceived by Mo Tzu two thousand years earlier. Among the philosophical innovations of the young Kant, I am intrigued by the immanent teleology already described. Other precritical insights impress me in their farsightedness too: that existence is not a property; that there are incongruent counterparts; that metaphysics must start with conceptual analysis; that there is no fundamental distinction between humans and other animals; or that biological diversity, the "*Mannigfaltigkeit der Natur,*" is intrinsically valuable. I find these ideas remarkable.

> **Uleman: There is something unsettling about reading about the young Kant trying to broker a peace between competing views. Do these facts about Kant's intellectual biography threaten to subordinate Kant's arguments to his own 'peace-maker' tendencies? Does this drain the life out of the arguments themselves, diverting them into biography, psychology, or historical contingency?**

Kant's peace-making tendencies are limited to his earliest work, the *True Estimation of Living Force* (1747). There he suggests a compromise between the two rival conceptions of force, hoping to make both Cartesians and Leibnizians happy. I agree with Professor Uleman that this drains the life out of the arguments, but I think this is no great loss, because these arguments are bad anyway. Apart from this, peace-making tendencies do not exist. To the extent that the precritical project involves a reconciliation, it is of a philosophical sort only: an attempt at harmonizing the then dominant paradigm of physical nature with generally accepted metaphysical desiderata of freedom, purpose, and God. Kant's argumentative strategies in the 1750s and 1760s show him to be indifferent to compromises. When Kant was impressed by an idea, he developed and applied it in innovative ways (as in his employment of Newton's lunar theory to the axial rotation); when he was not, he criticized it (as in his mockery of the anthropocentric fantasies of the Physico-Theologians). The aspiring thinker joined the philosophical debate with the words (the first sentence of his first book):

> I think I have reason to trust in the sense of the public enough that my freedom to contradict great men will not be regarded as a crime.[7]

> **Uleman: What do you think about the benefits and dangers of doing history of philosophy in general?**

The only danger of doing history of philosophy—otherwise a perfectly harmless enterprise—is to the historian herself: the risk of reducing one's thought to describing the creativity of others. The historian of philosophy is in the same predicament as the art critic or the literature scholar: doing the job well carries the risk of intellectual infertility and impotence. But this risk, I think, is outweighed by the benefits. The history of ideas teaches us about where we come from, and how we differ from others. It also beautifully illustrates an encouraging fact already recognized by Hegel: there *is* progress; things are getting better; and humankind is indeed evolving. Plus, there is the need to contribute to a record that is accurate and fair. To assume that historians of philosophy merely tread on well-trodden ground is false. Just consider Kant's own time, the 18th century, and compare the clichés with the facts. It is commonly assumed that Europeans developed the ideas of the Enlightenment then, and that not much happened in the time between Leibniz and Kant. The facts tell a different story: the ideas of the Enlightenment were not developed by Europeans per se, but were rather triggered by the failure of the Jesuit mission in China and its subsequent backwash of Confucianism pouring into Europe. The Leibnizian-Wolffian school philosophers, usually dismissed as uninteresting throwbacks and misguided metaphysicians, actually spearheaded the move towards secularization, towards a reconception of humans as citizens rather than subjects, and towards the universal validity of reason, regardless of background and gender. They popularized non-Western approaches, protected women thinkers in their ranks, and defended the freedom of thought against rabid Christians and other fundamentalists, often at great personal risk. Their story still needs to be told. There are still large white spots on our historical map.

Doing history of philosophy also promises personal benefits to the historian. Tracing the paths of the great minds is a superb schooling for learning the art and craft of philosophy. The historian accumulates not only data but also gains insight into creative strategies for innovation and discovery. I suspect that historians have better odds at becoming good thinkers than others. It's the same in the visual arts: a painter stands a greater chance at becoming a good artist if she learns how to draw first.

> **Uleman: What conclusions do you draw from this study—is the abandonment of the precritical project a failure? And is the critical philosophy itself, complete with transcendental idealism and distinct phenomenal and noumenal realms, also a failure—a brilliant one, to be sure, but a failure?**

The abandonment of the precritical project was *de facto* a failure for Kant, but one vastly outweighed by the insights of his critical philosophy. *That* was not a failure, on the contrary! I regard the critical philosophy as a great leap forward for human knowledge and progress. Of the three primary components of Kant's critical system, I would only shrug off his aesthetics as a contrived and spurious venture. But his other insights, on epistemology and ethics, are for me in a different league.

The Transcendental Dialectic of the *Critique of Pure Reason* strikes me as philosophy's definitive stance on the Great Metaphysical Questions of the universe, the soul, and God. Kant conclusively proved that any speculation on these questions will only amount to, well, speculation. Rational cosmology is dead. So is rational psychology. And we will never prove (or disprove) the claim of God's existence. All these are articles of faith, not matters of rational discourse. Kant's analysis of metaphysics is the final word on these riddles. We will never go beyond Kant.

The Transcendental Analytic of the first *Critique* has flaws, such as the artificial clockwork-model of the mind, or the futile search for the phantom of the synthetic *a priori*. The division between empirical appearances and an unknowable thing in itself is misguided too. Had Kant lived three generations later and written the *Critique of Pure Reason* after Darwin, he would have realized that the transcendental organization of sense impressions does not happen in thin air, but is the result of a successful evolution through environmental pressure and random mutations. We organize sensory information by means of certain cognitive forms instead of others because our forms *work*; if they didn't, *homo sapiens* as a species wouldn't have made it. That they work reveals a structural analogy between the way we pattern observations and how the things are patterned in themselves. But Kant lived before Darwin and therefore cannot be blamed for his ignorance. Aside from this, Kant's fundamental insight about the flows of data organization is simply correct. He settled the debate between rationalists and empiricists by showing that the mind is neither purely active nor merely passive but rather a combination of both. We now know that cognition is an interactive process that involves the ordering of sensory material by cognitive schemata, a give-and-play of reason and reality, and we owe this insight to Kant.

But perhaps Kant's most brilliant triumph is his practical philosophy. Certainly, the Categorical Imperative does not achieve everything that he hoped for. But it is the ultimate articulation of an ethical insight that has made the world a better place. Kant recognized the absolute worth of autonomous and rational beings, and the second version of the Categorical Imperative accordingly demands to treat people as ends and never just as means. Kant's gift to world civilization is the insight that autonomous beings are deserving of inviolable rights, and that it is more appropriate to socially organize them as empowered citizens than as disenfranchised subjects. The United Nations were first envisioned by Kant, and the Universal Declaration of Human Rights is inspired by his ethics. We are living in a better world because of these Kantian innovations, and we are reaping the benefits from his ethical and political insights. The Global Village is proving Kant right.

Notes

[1] Immanuel Kant, *Gedanken von der wahren Schätzung der lebendigen Kräfte* ("Thoughts on the True Estimation of Living Forces") written 1746-7; published 1749; Akademie ed. 1: 1-182.

[2] Martin Schönfeld, *The Philosophy of the Young Kant: The Precritical Project* (Oxford: Oxford UP, 2000) 205.

[3] Nelson Goodman, *Fact, Fiction, and Forecast*, orig. pub. 1955 (Indianapolis: Bobbs-Merril, 1973) 55.

[4] Sidney Axinn, "Kant and Goodman on Possible Individuals,"*The Monist* 61 (1978): 477-82.

[5] Cf. Immanuel Kant, *Gessammelte Schriften,* ed. Akademie der Wissenschaften (Berlin: Reimer, later DeGruyter, 1910 ff.) A218/B266, A230/B283; B288.

[6] Axinn 481.

[7] Immanuel Kant, *Lebendige Kräft* ("Living Forces"), preface #1, Akademie ed. 1:7, my translation.

Works Cited

Axinn, Sidney. "Kant and Goodman on Possible Individuals." *The Monist* 61 (1978): 477-82.

Goodman, Nelson. *Fact, Fiction, and Forecast.* 1955. Indianapolis: Bobbs-Merrill, 1973.

Kant, Immanuel. *Gessammelte Schriften.* Ed. Akademie der Wissenschaften. Berlin: Reimer, later DeGruyter, 1910ff.

Kant, Immanuel. *Gedanken von der wahren Schätzung der lebendigen Kräfte* ("Thoughts on the True Estimation of Living Forces.") Academie ed. 1749.

Schönfeld, Martin. *The Philosophy of the Young Kant: The Precritical Project.* Oxford: Oxford UP, 2000.

.

Three Levels of Self-Deception

Critical Commentary on Alfred Mele's
Self-Deception Unmasked (Princeton and Oxford: Princeton UP, 2001)

Peter Dalton, *Florida State University*

In *Self-Deception Unmasked*, Alfred Mele focuses almost entirely on explaining self-deception, and he argues effectively for explanations that stress non-agency (and hence that are causal in nature) rather than ones that cite agency (and hence that involve intention, trying and purpose). I want to do something that may seem irrelevant to his concerns, for I want to focus on the conceptual analysis of self-deception. I hope to show, however, that the proper analysis of self-deception points toward the general correctness of explanations stressing non-agency.

I believe that an analysis of self-deception must rest on a recognition of three epistemic levels. If we let 'W' stand for a belief, then the three levels that pertain to self-deception are as follows:

(1) A person believes W, and W is false.

(2) This person believes W because of some incorrect thinking.

(3) This person is not aware of the incorrect thinking that has led him to believe W,
and he either believes that he has thought correctly or at least does not believe that
he has thought incorrectly.

On my view, (1) is the level of *falsehood*, (2) of *deception*, and (3) of *self-deception.* Mele's book does not concern itself with (1), nor should it; its focus is psychology, or more generally the workings of the human mind. His book stays almost entirely at level (2), as nearly every page discusses varying attempts to categorize or explain the kinds of incorrect thinking that lead to false beliefs at level (1). The question is whether his, or any other, explanation of self-deception can restrict itself to level (2). I doubt it. I think every such account must proceed eventually to level (3), for reasons I'll now recount.

As Mele recognizes, while every case where a person believes a falsehood may be labeled "deception," not every such deception is a case of self-deception. Is it certain kinds of level (1) beliefs that mark out a case of deception as self-deception? Someone might think so, as the phrase "self-deception" seems to imply some kind of deeply flawed belief about one's self. This is dubious, however, since many of the beliefs held by self-deceived people aren't about themselves; and while all such beliefs in a broad sense concern the believer, that alone doesn't mean that the believer is self-deceived. It is possible that any belief could, in the right context, involve a person in self-deception (e.g., a philosopher might become so obsessed with the arguments of the *Meditations* I, that she comes to believe that 2 plus 3 might not equal 5.)

Is it certain kinds of incorrect thinking that mark a case of deception as one of self-deception? If this were true, we could analyze self-deception by sticking to level (2). Someone might get the impression that Mele believes this since his sufficient conditions for self-deception might seem to limit themselves to levels (1) [criterion (1)] and (2) [criteria (2)-4)]:

1. The belief that *p* which S acquires is false.
2. S treats data relevant, or at least seemingly relevant, to the truth value of *p* in a motivationally biased way.
3. This biased treatment is a nondeviant cause of S's acquiring the belief that *p*.
4. The body of data possessed by S at the time provides greater warrant for ~*p* than for *p*.[1]

But since Mele doesn't require that everything that leads someone to believe "*p*" (his symbol for the false belief) occur at level (2), he may be open to an analysis of self-deception that moves on to level (3).

Here's why I think we need to move to level (3). Someone who is self-deceived is ignorant in an important way about something that concerns himself. The problem isn't simply that he doesn't know something about himself, it's the kind of thing he doesn't know and why he doesn't know this. Think of Sartre's classic examples of bad faith: the woman who can't decide if a man is making a sexual advance toward her, the waiter who confuses his put-on waiterly role with his genuine self, and the man who wonders if his homosexual acts make him a homosexual.[2] Each doesn't know something about himself or herself. Each seems to hold some false beliefs about herself or himself. The problem, which Sartre subtly depicts, lies in the confused and illogical thinking that leads these people to hold these false beliefs. It isn't just that these people think incorrectly; if it were, their self-deception would be confined to level (2). The real problem—what most disturbs the reader—is that they don't know that they think this way, and the interesting question is why they don't know this. Each has some dim awareness that their thinking may be amiss, but it's part of their self-deception that they can't quite figure this out. If they could, it's unlikely that they would believe W. This is why I think that a proper analysis of self-deception must involve level (3). Deception about one's self involves a lack of knowledge of the incorrect thinking that leads one to be deceived about something else. This is why self-deception can center on a belief that isn't about oneself (e.g., the belief that one's wife is having an affair). But that's the deception, a level (2) matter. To get self-deception, we must bring in a self that is ignorant, confused or thinking wrongly at level (3) about the thinking that has led it to be deceived at level (2). Mele's book uses the apt metaphor of a mask. If self-deception exists at both levels (2) and (3), then it involves a two-sided mask, one that hides both the falsity of W and why one came to believe W.

Does this mean that Mele's book is one big category mistake, since he confines himself to level (2)? No. My hunch is that the kinds of psychological errors, flaws and habits he cites as explaining level (2) deception will also help us explain level (3) self-deception (e.g., people who have fragile self-esteem are prone to err at all levels, and a man who would be extremely upset if he discovered that his wife has been unfaithful to him will not critically reflect on the thinking that has led him to believe that she has always been faithful to him.). I doubt that Mele's explanations can provide the whole story about level (3) self-deception. While they focus on the incorrect thinking that led a person to believe something that is false, I don't think they focus enough on why the person is unaware of the incorrectness of this thinking.

Now for a word about agency. We're often blind to self-deception. Indeed, I think a proper analysis requires some such blindness. Self-deception also calls for thinking. While some philosophers characterize thinking in a way that makes it an act or at least a result of agency, these conceptions are too narrow. Sometimes thinking just occurs. Sometimes thinking is uncontrolled (e.g., the way background information influences an inference). And sometimes thinking is forced on us (e.g., obsessive thinking). Lastly, self-deceived thinking is incorrect. Mele persuasively argues that incorrectness is usually not intended or sought, which means it's not the result of agency or at least not the result of responsible, knowing agency. If all this is right, if self-deception is not the work of agency but is something we undergo, then we have a strong reason for agreeing with Mele that explanations of self-deception must stress non-agency.

Notes

[1] Alfred Mele, *Self-Deception Unmasked* (Princeton and Oxford: Princeton UP, 2001) 50-51, 120.

[2] Jean Paul Sartre, *Being and Nothingness,* trans. Hazel Barnes (New York: Washington Square Press, 1956) 96-98, 101-103, 107-108.

Works Cited

Mele, Alfred. *Self-Deception Unmasked.* Princeton and Oxford: Princeton UP, 2001.

Sartre, Jean Paul. *Being and Nothingness.* Trans. Hazel Barnes. New York: Washington Square Press, 1956.

Capturing Our Attitude Toward the Self-Deceived

Critical Commentary on Alfred Mele's
Self-Deception Unmasked (Princeton and Oxford: Princeton UP, 2001)

Crystal Thorpe, *University of Florida*

In the book, *Self-Deception Unmasked*, Alfred Mele offers a fresh new approach for dealing with the seemingly intractable problem of self-deception. Traditionally, philosophers have taken self-deception to be a phenomenon in which an individual who initially knows or believes the truth, *p*, intentionally causes herself to believe *~p*.[1] Two things are important to note about the traditional view. First, self-deception is taken to be an intentional process; and second, the self-deceived agent is characterized as holding conflicting beliefs, that is, she holds both her initial belief that *p* and her newly acquired belief that *~p*. Mele rejects the traditional view by arguing that garden-variety cases of self-deception are not intentional and do not involve the holding of contradictory beliefs. His main ground for rejecting the traditional view is that it does not adequately explain what happens in garden-variety cases of self-deception.

Mele argues that we have evidence from empirical psychology that suggests that typical cases of self-deception are not intentional and do not involve the holding of contradictory beliefs.[2] I think Mele is right to reject the traditional view. The reason I think the traditional view should be rejected, however, has less to do with evidence we get from empirical psychology and more to do with our attitudes toward self-deceived people. The traditional view fails to capture our attitudes toward self-deceived people. If the traditional view were correct, one would expect self-deceived people to be regarded as desperate, epistemically irresponsible, irrational and perhaps even mentally ill. Yet considering the prevailing social attitudes, how do we customarily regard the broker who thinks she is easier to get along with than all the other associates, or the wife who is self-deceived in believing her husband is not having an affair, or the parent who is self-deceived in believing that her child has not committed a felony? Even if we tend to perceive such individuals to be desperate, epistemically irresponsible, irrational or mentally ill, surely we do not view the majority of self-deceived people in this way. Rather, we tend to regard self-deceived people as being irritating, pitiable, silly or laughable. The prevailing attitudes are compatible with Mele's model of self-deception and incompatible with the traditional model.

I take it that to be self-deceived is a bad thing. It is not something for which one should strive. Quite to the contrary, it is something that one ought to avoid. If one takes the traditional view of self-deception, it is easy to see why it is bad. For starters, the self-deceived person violates several epistemic norms. A self-deceiver starts off with the belief that *p*, and then causes herself to have a belief that *~p*. She ends up, therefore, with the belief that *p* and the belief that *~p*. This clearly violates epistemic norms governing the consistency of our beliefs. Furthermore, self-deceivers violate norms governing how evidence is to be gathered and interpreted. Now, if the individual violates these norms unintentionally, we are not likely to attribute epistemic irresponsibility to her. The traditionally conceived self-deceiver, however, violates these norms *intentionally*. To make matters worse, the reasoning that leads the self-deceiver to violate these norms is itself irrational.

Let me say a bit more about what I mean here. Assume that A wants a certain proposition, *p*, which she believes to be false, to be true. For example, she may believe that the proposition, "My spouse is not having an affair," is false, yet she may want it to be true. Given that she believes that her spouse *is* having an affair, she presumably has evidence that supports this belief. For example, her spouse arrives home late, smelling of perfume, at least two nights a week. Her spouse takes two hour lunch breaks out of the office when in the past he routinely ate at his desk. Her spouse has been seen with another woman in intimate settings. And so on. Now, the thought that her husband is having an affair is devastating to our agent. If she thinks about it, she finds herself unable to eat, sleep or function. She desperately wants it to be the case that he is not having an affair. What can she do about this? On the one hand, she could break up the relationship. Notice that if she were to do this, the proposition "My spouse is not having an affair" would be true. This, however, will not solve her problems, for not only does she want the proposition "My spouse is not having an affair" to be true, she wants the proposition "My spouse *didn't* have an affair" to be true as well. So, in a desperate attempt to gain control over a situation that is completely out of her control, she decides to cause herself to believe that her husband is not having an affair.

Now, intentional action is made up of three components. One, the agent sets a goal for herself. Here the goal is to believe something she knows to be false. Two, the agent figures out, through a process of rational deliberation, the means to that goal. Here, the means involve deliberating badly at just the right points. And three, the agent acts on this means in order to further that goal. That is, she deliberates badly in just the right way to cause herself to believe that her husband did not have an affair. Clearly, there is something wrong here. Although this intentional act may be successful, that is, although our agent may succeed in causing herself to believe that her husband is not having an affair, and although there may be something rational about the strategy *itself*, the thinking that led her to perform this intentional act *is itself irrational.*

If the traditional view were correct, and the self-deceived agent behaved as I have described above, we would tend to think her desperate and epistemically irresponsible, not to mention irrational. We would find her desperate, for only a desperate individual would intentionally engage in such irrational behavior. We would see her as being epistemically irresponsible as well, for it is epistemically irresponsible to intentionally flout epistemic norms. Furthermore, we would perceive her to be irrational, for intentionally causing oneself to believe something that is false is clearly irrational. It is no help to say that the traditionally conceived self-deceived agent *unconsciously* intends to deceive herself or that there is some mental partitioning that takes place. Hypothesizing such mechanisms does help to explain how self-deception, traditionally conceived, is psychologically possible. Furthermore, it takes some of the sting out of its irrationality. However, if such mechanisms were indeed at work in the self-deceived person, most of us would tend to think that the self-deceived person verges on being mentally ill.

I realize that my portrayal of the traditionally conceived self-deceived person may be a bit unfair. However, if you take seriously the claim that self-deception is an intentional phenomenon, the self-deceived person becomes one to whom you would attribute desperation, epistemic irresponsibility, and worst of all perhaps even mental illness. Mele rejects the claim that self-deception is intentional and in so doing he presents us with a characterization of the self-deceived person that more closely matches our attitudes towards the self-deceived. On Mele's view, an agent does not intentionally cause herself to believe something that is not the case. Rather, cognitive mechanisms are activated in us by strong feelings and desires that have the effect of making us believe something that is not true. Often these mechanisms are activated without our being aware of it. Sometimes, on the other hand, we intentionally activate them. In these cases, however, we don't activate them in order to deceive ourselves. Rather, we have another goal in mind. Perhaps we do it to avoid pain or to make ourselves feel happy—as Mele's character Beth intentionally focuses on pleasant memories of her father in order to avoid the painful truth that he did not love her best.[3] Although Beth's goal is to avoid pain, she inadvertently causes herself to believe that her father cared for her more than he actually did.

On Mele's account, self-deception is something that happens to us rather than something that we do. This isn't to say, however, that we cannot avoid it. One can reflect on one's motivations and cognitive processes and try to stop oneself from becoming self-deceived. Notice, however, that when one fails to do this, we do not tend to attribute epistemic irresponsibility to that person. Rather, we tend to pity that person or become irritated with her. Furthermore, there is nothing desperate about Mele's self deceived person. This, in part, is because the phenomenon as Mele sees it is all too human. It happens to the best of us. Mele's self-deceived agent captures our attitude towards the self-deceived. The traditional view does not.

Notes

[1] See D. Davidson, "Deception and Division" in E. LePore and B. McLaughlin, eds., *Actions and Events* (Oxford: Basil Blackwell, 1985) 138-48 and D. Pears, *Motivated Irrationality* (Oxford: Oxford UP, 1984).

[2] See Chapter 2 of Alfred Mele's *Self-Deception Unmasked.* Princeton and Oxford: Princeton UP, 2001.

[3] Mele 18-19.

Works Cited

Davidson, D. "Deception and Division." *Actions and Events.* Eds. E. LePore and B. McLaughlin. Oxford: Basil Blackwell, 1985. 138-48.

Mele, Alfred. *Self-Deception Unmasked.* Princeton and Oxford: Princeton UP, 2001.

Pears, D. *Motivated Irrationality.* Oxford: Oxford UP, 1984.

Some Remarks on Self-Deception: Mele, Moore, and Lakatos[1]

Critical Commentary on Alfred Mele's
Self-Deception Unmasked (Princeton and Oxford: Princeton UP, 2001)

***Risto Hilpinen**, University of Miami*

Sartre and Moore on Contradicting Oneself

Much of the recent philosophical discussion of the problems and paradoxes of self-deception or self-delusion goes back Jean-Paul Sartre's analysis of bad faith in *Being and Nothingness*: many recent papers on the subject introduce the concept of self-deception by references to Sartre. However, people have been writing about the puzzles of self-deception for centuries. The earliest book on the subject I found in the library of my university was published in 1614, Daniel Dyke's *The Mystery of Selfe-Deceiuing. Or A Discourse and Discouery of the Deceitfulnesse of Mans Heart.*

According to Sartre, bad faith consists essentially in lying to oneself. [2] The possibility of lying depends on the ontological and epistemic duality between the deceiver and the deceived, but how can this duality be preserved if the two parties are in the same consciousness, that is, if the deceiver is trying to hide the truth from himself? This is a good conceptual puzzle for philosophers to write about. In Sartre's formulation, the philosophical problem of self-deception seems to be the question: How is it possible to lie to oneself, that is, *qua* the deceiver accept a proposition, and *qua* the deceived party not accept the same proposition, or even accept the contradictory proposition?[3] According to this model, self-deception involves the acceptance of mutually contradictory propositions, and consequently the problem of self-deception is a special case of the more general question about the possibility of having jointly inconsistent beliefs. Makinson's paradox of the preface[4] and the lottery paradox are standard examples of inconsistent belief sets. I accept all my beliefs, but on the basis of our general fallibility, I also believe that some of my beliefs are false.[5] This is a logically inconsistent system of beliefs. In the same way, in the case of a fair lottery in which just one of a large number of tickets will win a prize, it seems reasonable to believe about each ticket x that x will not win, and also believe that one of the tickets will win.[6] These examples are not examples of self-deception and seem to have very little to do with self-deception. If self-deception is analogous to lying to another person, it involves inconsistency in a particularly acute form, namely, the acceptance of two mutually contradictory propositions and not just the acceptance of an inconsistent system of beliefs. But not all inconsistencies are instances of self-deception. Even if it were possible to believe explicitly contradictory propositions, such a situation could not be regarded as an example of self-deception unless one of the contradictory beliefs were in some way *hidden* from the believer. According to Harold Sackeim and Ruben Gur, this means that the individual is not aware of holding one of the beliefs, in other words, a person who is deceiving himself about *p* accepts (believes) both *p* and ~*p*, but does not believe that he believes that *p* or does not believe that he believes that ~*p*.[7] In his new book, *Self-Deception Unmasked* (as well as in his earlier publications), Alfred Mele rejects this view.[8] The view that self-deception must involve contradictory beliefs involves a common philosophical misconception, namely, that one can contradict oneself only by accepting contradictory propositions or at least a set of propositions which are jointly inconsistent. There are many well-known counterexamples to this view, even though they have not always been recognized as such. One such example is the paradoxical assertion (type) considered by G. E. Moore:

It is raining, but I do not believe that it is raining.[9]

This assertion has the form

(AsMoore) R & not Bel(I, R).

This proposition is not self-contradictory, but by uttering it the speaker contradicts himself. The assertion of a conjunctive proposition is a conjunctive assertion. Thus

(AsM) As(R & not Bel(I, R)), where 'AsP' means that the speaker asserts that P,

entails

(1.1) As(R)

and

(1.2) As(not Bel(I, R)

A sincere assertion is an expression of belief: by asserting a proposition *p* the utterer *states* that *p* and *expresses* the belief that *p* (*conveys* the information that she believes that *p*); thus the former assertion (1.1) conveys the information that I believe that R, and in the latter assertion (1.2) I *assert* that I do not believe that R.[10] The two assertions cannot be "correct" at the same time: either the first is not sincere or the second is false. This can be regarded as an instance of "contradicting oneself." In this sense anyone who utters a Moore sentence contradicts himself. The indefensibility (or inconsistency) of an assertive utterance of a Moore sentence can be explained without assuming that the speaker utters a contradictory proposition.[11] Even though an utterer of a Moore sentence does not assert contradictory propositions, she is not "unanimous," as Mrs. Slocombe would put it.

In the same way, a person who believes that R, but thinks (believes) that she does not believe that R, does not necessarily accept contradictory propositions. (I assume here that believing that R does not entail and is not entailed by believing that one believes that R.) However, such a person could not express or articulate these beliefs without contradicting herself in the sense described above. If she were to make the conjunctive assertion that (i) she believes that R, and (ii) she believes that she does not believe that R, she would *express* by means of the first conjunct that that she believes that she believes that R, and simultaneously *state* by means of (ii) that she believes that she does not believe that R. A person who is in this way mistaken or deceived about her beliefs is subject to an interesting form of self-deception: she is deceived, even though no one else is deceiving her. This conception of self-deception differs from Sackeim and Gur's conception: according to Sackeim and Gur, a self-deceiver believes *p* and~*p*, but does not believe that he believes (for example) *p*, whereas a Moorean self-deceiver simply believes that *p* and also believes that he does not believe *p*.

Self-Deception and Motivational Bias

Many philosophers have approached the phenomenon of self-deception by starting from certain representative examples and not from an abstract model. Such examples are not hard to come by. Here is one: *The Miami Herald*, a Miami newspaper, reported on October 26, 2001, that Senator Bob Graham had stated:

> I am confident that we will eventually achieve our objective of not only taking down bin Laden but other global terrorists, because it is too important for us not to be successful.[12]

According to a newspaper report published on October 25, Defense Secretary Donald Rumsfeld had expressed a less optimistic view. He had said: "I just don't know whether we will be successful" at tracking bin Laden down.[13] In view of Secretary Rumsfeld's observation, Senator Graham's statement may look like an instance of self-deception, an expression of a self-deceptive belief. However, this depends on how the statement is understood. It can be taken to mean that Senator Graham's confidence (and belief) that bin Laden will eventually be caught is directly caused by (and dependent on) his view that it is important to be able to do it. If the statement is understood in this way, it can be regarded as an example of self-deception. Alternatively, the statement can be taken to mean that the importance of catching the terrorists is *evidence* that they will be caught: the importance of the goal is evidence that the attempt to catch the malefactors will not be given up until they have been caught. According to this (more charitable) interpretation, the statement does not involve any self-deception.

One of the standard examples discussed in the philosophical literature is the following. A husband (let us call him Adam) believes that his wife Eve has always been faithful to him despite strong evidence to the contrary. For example, let us assume the couple has never had any sexual relations (because the husband is impotent or for some religious reason), but Eve is about five months pregnant. When Adam questions Eve about the situation, she assures him that she has been completely faithful and that her pregnancy is a miracle. The fact that the scarcity of hotel rooms forced her to share a room with a male co-worker during a business trip about five months ago has nothing to do with the matter. Adam accepts this explanation and believes her. This example fits Alfred Mele's analysis of self-deception (given the obvious assumption that Eve has in fact been unfaithful). According to Mele, the following conditions are jointly sufficient for S's self-deception in acquiring a belief that *p*:[14]

(2.1) The belief that *p* which S acquires is false.

(2.2) S treats data relevant, or at least seemingly relevant, to the truth-value of *p* in a motivationally biased way.

(2.3) The biased treatment is a nondeviant cause of S's acquiring the belief that *p*.

(2.4) The body of data possessed by S at the time provides greater warrant for ~*p* than for *p*.

Mele's conditions concern the acquisition of self-deceptive beliefs. They can be applied to belief retention by replacing the expression "acquiring" in condition (2.3) by the expression "sustaining":[15]

(2.3') The biased treatment is a nondeviant cause of S's sustaining the belief that *p*.

Conditions (2.3) and (2.3') are causal conditions. A cause-effect relation between two events or states involves causal dependence or a chain of causal dependence relations between them.[16] In ordinary cases of self-deception, S's belief that *p* is caused and causally dependent on his biased treatment of the evidential data; thus (2.3) entails the following dependence condition:

(2.5) If S did not treat the data in an evidentially biased way, S would not believe (or would not have acquired the belief) that *p*.

Moreover, in ordinary cases of self-deception, for example, in the case of Adam and Eve, the evidential bias as well as the belief that *p* are dependent on S's interest in the truth of *p* (S's preference for *p* over ~*p*). If S did not treat the data in a biased manner, S would not be able to sustain the self-deceptive belief, and S treats the data in a biased manner because S prefers the truth of *p* to its falsity (in the sense of wishing *p* rather than ~*p* to be true). If Eve's faithfulness were not important to Adam, he would not treat the evidence in a biased way and would not continue to believe that Eve is faithful despite the apparently conclusive counter-evidence. These dependence relations are expressed by the following conditionals:

(2.6) If S did not prefer *p* to ~*p*, S would not treat the evidential data in a biased way, and

(2.7) If S did not prefer *p* to~*p*, S would not believe that *p*.

(2.6) expresses part of Mele's condition that S treats the evidence in a "motivationally" biased way.

Mele's condition (2.3) contains the poorly understood expression "nondeviant cause" often encountered in the literature on the philosophy of action; this qualification distinguishes (2.3) from the simple dependence condition (2.5). Mele gives the following example (adapted from Robert Audi) of a situation in which a false belief is caused by the believer's evidential bias and condition

(2.5) holds, but no self-deception occurs because S's belief does not depend on his evidential bias in the right way.[17] Bob is investigating an airplane crash, and hopes that it was the result of a mechanical failure rather than a terrorist attack. He consults Eva, who has usually rejected terrorist hypotheses in the past, but in this instance Eva believes (on the basis of her data) that the terrorists were at work, and is able to convince Bob that the terrorist hypothesis is true. However, the crash was not caused by any terrorists, but by a mechanical failure, and this is clearly shown by the evidence possessed by the investigators (other than Eva), and readily accessible to Bob. We can assume that without his bias in favor of the mechanical failure explanation Bob would not have limited his inquiries to Eva, and Eva's erroneous opinion would not have led Bob to accept the false terrorist hypothesis.

For the purpose of this example, the warrant condition (2.4) must be (re)interpreted in such a way that the evidence which provides greater warrant for *~p* than *p* is evidence "readily accessible" or *available* to S rather than the evidence actually possessed (accepted or known) by S. Bob could hardly accept the terrorist hypothesis rather than the mechanical failure explanation (which he prefers) if he were fully aware of the evidence which supports the latter hypothesis. According to Mele, this is a permissible interpretation of the warrant condition.[18]

This example has the following form: Assume that Bob wishes *p* to be false, and this leads him to search evidence relevant to *p* in a biased way (to prove the falsity of *p*), but to his great surprise most of the evidence turns out to be favorable to *p*. On the basis of this evidence, Bob cannot but believe that *p*. Even if *p* turns out to be false, Bob is not subject to self-deception. However, Bob's believing that *p* is dependent on his biased evidence gathering procedures: we can assume that without his bias he would have conducted the investigation into the truth of *p* in an impartial manner, and would have formed his opinion in accordance with the readily available evidence against *p*. Thus condition (2.5) can hold (together with (2.1)-(2.2) and (2.4)) also in cases in which no self-deception occurs.

According to Mele, a selective approach to gathering evidence for a proposition *p* owing to a desire that *p* can contribute to self-deception by "leading one to overlook relatively easily obtainable evidence for *~p* while finding less accessible evidence for *p*, thereby leading to believe that *p*."[19] This does not hold in the present example: Bob's approach,

> an instance of selectively gathering evidence for P motivated by a desire that P—is of a kind that leads to self-deception by increasing the subjective probability of the proposition that the agent desires to be true, not by increasing the subjective probability of the negation of that proposition.[20]

Bob's attempt to find evidence for the hypothesis he wishes to be true causes him to find evidence against the hypothesis and leads him to conclude that the hypothesis is false. For this reason, Mele does not regard this example as an example of self-deception, and observes that "S enters self-deception in acquiring the belief that *p* if and only if *p* is false and S acquires the belief in 'a suitably biased way'."[21] The purpose of the condition of non-deviance in (2.3) is to register that the evidential bias must be "suitable" for self-deception. Bob's belief that the crash was caused by a terrorist attack is not an instance of self-deception because Bob initially preferred the explanation by mechanical failure, and tried to find evidence supporting the latter explanation, but ended up accepting the terrorist hypothesis. Bob's belief was not consonant with his interests or desires. This means that the dependence condition (2.7) does not hold, and it might be suggested that the "deviance" of the dependence of Bob's belief on his evidential bias is due to the failure of (2.7). In proper cases of self-deception, an agent's belief that *p* should depend not only on his evidential bias in favor of *p*, but also on his preference for *p* over *~p* (and not on his preference for *~p* over *p*). It is clear that Mele makes this assumption in his discussion of the case.

However, the dependence condition (2.7) cannot be regarded as necessary for self-deception if we accept the possibility of "twisted" instances of self-deception in which a person's false belief that *p* is dependent on his desire that *~p*, that is, on his preference for *~p* over *p*. An "irrational" (i.e., unfounded) false belief that *p* can be caused by the fear that *p* rather than the desire that *p*; in such a case an emotion (for example, jealousy or fear) leads a person to form or retain "an intrinsically unpleasant belief against the promptings of reason."[22] In "twisted" cases, we seem to have, instead of (2.7),

(2.8) If S did not prefer *~p* to *p*, S would not believe that *p*.

If a jealous husband did not wish his wife to be faithful to him, he would not believe that his wife is unfaithful. The following simple formula covers both forms of dependence:

(2.9) S's belief that *p* depends on the desirability of *p* (for S).

However, as the example about Bob and Eve shows, this condition is not always sufficient. As Mele observes, a conceptually satisfactory account of self-deception must say more about the "routes" to self-deception, that is, about the way in which an agent's belief depends on his motivation and his evidential bias.[23]

According to Mele's first condition, self-deception involves a false belief. He presents conditions (2.1)-(2.4) as jointly sufficient for self-deception, but does not regard them as necessary conditions. However, he takes (2.1) to be a necessary condition of self-deception. He says that the first condition "captures a purely lexical point: a person is, by definition, *deceived in* believing that *p* only if *p* is false; the same is true of being *self-deceived in* believing that *p*."[24] Mele is considering self-deception *in believing* something, and condition (2.1) holds for this concept by definition, but it is interesting to observe that withholding judgment (agnosticism) about some proposition can look very much like self-deception and can be motivated in the same way. For example, if Adam simply refuses to believe that Eve is unfaithful (without claiming that she is faithful), and continues to insist (against overwhelming evidence) that he has no idea whether Eve has been unfaithful or not, he seems to be engaged in a form of self-deception. Adam acts like a jury which fails to convict a defendant despite practically conclusive DNA-evidence against him. (A "not guilty" verdict is not an assertion that defendant is innocent, only that the evidence is not regarded as sufficient to prove that he is guilty.) In a situation of this kind, Adam's self-deception need not involve the acceptance of any false belief, but consists in having an incorrect and self-deceptive *attitude* towards a proposition. This is self-deception *in refusing to believe* what should be (and perhaps is) obvious to any reasonable person.

Lakatos, Confirmation Bias, and Self-Deception

The example about Adam and Eve reminds me of Imre Lakatos's conception of the methodology of scientific research programs. According to Lakatos, a scientific research program has three components: (i) a "hard core" of theoretical laws, together with a "protective belt" of auxiliary hypotheses which can be used to explain away apparent counter-evidence to the theory; (ii) the negative heuristic, that is, methodological rules which prohibit the application of *modus tollens* to the hard core of the program; and (iii) the positive heuristic of the program which gives directions for future development and for possibly fruitful auxiliary (protective) hypotheses.[25] According to Lakatos, research programs can be either "progressive" or "degenerative." A progressive program is capable of using its positive heuristic successfully to predict novel phenomena, whereas a degenerative program can account for anomalous phenomena only by inventing auxiliary hypotheses after such phenomena have been discovered.

In the example given above, Adam's conviction that Eve is faithful is part of the hard core of his conception of their marriage, and he explains apparent counter-evidence by introducing auxiliary hypotheses (such as the miracle hypothesis) for its protection. The "research program" by which Adam sustains his belief seems degenerative insofar as the auxiliary hypotheses introduced for the purpose of protecting the core belief (the miracle hypothesis or, for example, the hypothesis that when Eve had her annual checkup about 5 months ago, she was artificially inseminated by mistake) do not lead to successful predictions, but must be protected by additional auxiliary hypotheses. From a Lakatosian perspective, self-deception means clinging to a degenerative belief revision program built around a false core hypothesis. We might say that scientists and philosophers who cling to degenerative research programs in an obsessive manner are engaged in a form of self-deception.

The Lakatosian model may also throw some light on what has sometimes been called "confirmation bias" or "verification bias," people's tendency to focus on evidential information that confirms and avoid or overlook evidence which disconfirms their current beliefs and hypotheses.[26] Thus, according to the confirmation bias thesis, a methodologically unsophisticated person who is testing a hypothesis tends to focus on the confirming evidence rather than disconfirming evidence. People are usually not good Popperians.

According to Mele, the confirmation bias contributes to the evidential bias which is one of the conceptual ingredients of self-deception:

> Given the tendency that this bias [the confirmation bias] constitutes, a desire that *p*—for example, that one's child is not experimenting with drugs—may, depending on one's desires at the time and the quality of one's evidence, promote the acquisition or retention of a biased belief that *p* by leading one to test the hypothesis that *p*, as opposed to the hypothesis that ~*p*, and sustaining such a test.[27]

This is puzzling, because from the logical point of view there is no difference between testing *p* and testing ~*p*: any test of a hypothesis *p* is simultaneously a test of its negation. A test of a hypothesis *p* is an attempt to find an answer to the question whether *p* is true, and the following three sentences express the same question:

(i) Is *p* true (or false)?
(ii) Is ~*p* true (or false)?
(iii) Is *p* true or is ~*p* true?

It should not make any difference whether a test is described as a test of *p* or as a test of ~*p*. In the discussion and interpretation of psychological experiments, it is important to distinguish an investigator's (a psychologist's) theoretical language or "system language" from the language of the subjects who are being investigated. The instructions given to the subject at the beginning of an experiment belong to the latter. It is perfectly possible that in an experiment about reasoning, a subject's interpretation of the evidential data depends on the way in which the instructions are formulated. If a mother were asked to "test" the hypothesis that her daughter *is* experimenting with drugs, her test procedures and conclusion might differ from those prompted by the instruction to determine whether her daughter is *not* experimenting with drugs. This is possible, but it would be surprising: if the mother has a tendency to deceive herself, that is, if her interpretation of the evidence depends on what she wishes to be true, she is in both situations likely to overlook evidence which would be positively relevant to the drug hypothesis. If the mother is thought to be testing the drug hypothesis, she is likely to show a "disconfirmation bias." This is, of course, an empirical issue, to be decided by means of experiments.

The alleged confirmation bias is related to, and difficult to distinguish from, a number of other "biases of rationality" studied by psychologists.[28] According to Fischoff and Beyth-Marom,

> Confirmation bias has proven to be a catch-all phrase incorporating biases in both information search and interpretations. Because of its excess and conflicting meanings the term might be retired.[29]

However, the term has not disappeared from the psychological literature. The clearest instances of confirmation bias can be found in situations in which a subject is looking for evidence relevant to a proposition he already believes (or thinks he knows to be true). Evans and Over have distinguished confirmation bias—the tendency to seek evidence that supports a prior belief—from *belief bias*, a "biased evaluation of the evidence that is encountered."[30] It is clear that both are involved in the ordinary cases of self-deception. In rational belief revision, the evaluation and interpretation of new evidence depends usually on the believer's prior beliefs and on their degree of (epistemic) *entrenchment* in her belief system.[31] It might be suggested that, in cases of self-deception, the dependence of the evaluation and interpretation of new evidence on the agent's prior beliefs depends on her interests and desires; thus self-deception seems to involve a second-order dependence relation. Another form of bias is "positivity bias," the tendency to favor and find confirmation for hypotheses expressed in positive terms (instead of negative terms).[32] This bias seems to have been shown by the subjects in an experiment reported by Trope, Gervey and Liberman:

> Subjects who tested the hypothesis that a person was angry interpreted that person's facial expression as conveying anger, whereas subjects who tested the hypothesis that the person was happy interpreted the same expression as conveying happiness.[33]

"X is angry" and "X is happy" are not negations of each other, but the experimenters presumably regarded them as incompatible descriptions. The first-mentioned subjects were trying to answer the question whether a person shown to them was angry or not, whereas the subjects of the second experiment were trying to find out whether the person shown was happy or not. The results of the experiment illustrate the positivity bias rather than the belief bias or the confirmation bias.

The direction of the "confirmation bias" seems to depend on how a given test is described. Yaacov Trope and Akiva Liberman provide an answer to this puzzle: not every proposition counts as a hypothesis. The negation of a hypothesis need not be a hypothesis (of the proper kind). Trope and Liberman observe:

> Any given hypothesis is usually more specific than its alternatives. A hypothesis often refers to a single possibility (e.g., the target is a lawyer), whereas the alternatives may include a large number of possibilities (e.g., the target has some other occupation).[34]

When Trope and Liberman refer to the "alternatives" of a given hypothesis, they seem to mean its negation, that is, the disjunction of all its alternatives (in their example, the disjunction of the occupations other than a lawyer). A hypothesis that refers to a "single possibility" is obviously more informative and has more explanatory value than its negation. For example, the hypothesis "Dr. Kafka is a professor" is a good and informative answer to a question about Dr. Kafka's profession, but its negation is almost worthless. A specific and informative hypothesis is attractive not only if a person wishes it to be true, but also on the basis of its informational value. Both the acceptance and the rejection (the acceptance of the negation) of such a hypothesis adds information to a person's belief system, but the acceptance of the hypothesis adds more information than its rejection. This may lead to a form of "rational" confirmation bias (or positivity bias), based on the believer's epistemic interest in having informative beliefs, and should not be confused with other forms of self-deception. A highly informative hypothesis can function in the same way as the "hard core" of a Lakatosian research program: once accepted, an investigator is reluctant to give it up (let alone reject it) unless it can be replaced by an equally informative alternative. According to Lakatos, the hard core of a research program is protected by maximal "confirmation bias": the negative heuristic of the program instructs the investigator to protect it under all circumstances by suitable auxiliary hypotheses, and never to abandon it.

Concluding Remarks

To conclude, I would like to suggest in what sense a self-deceiver can be said to contradict himself without accepting contradictory propositions. In philosophical discussion believing something (or the belief that *p*) is often construed simply as an attitude towards a proposition (the acceptance of a proposition) or as having a proposition in one's mental "belief box." I think belief (or believing) is more complex and some puzzles of self-deception are due to this complexity. Believing something seems to involve several conceptual constituents. I would like to suggest that a full-fledged belief (for example, the belief that it is raining) involves the following:

(B1) Assent to the proposition and a disposition to assert (utter) the proposition in appropriate circumstances. The assent may be external (linguistic) assent, or merely internal, mental assent.

(B2) Disposition to act in a way that would be optimal (given the believer's interests) if the belief were true.

(B3) A conception of what the world would be like if the belief were true, which involves knowing how to find out whether the belief is true and how to defend the belief against objections.

The first condition may be termed the assent condition (As-condition), and (B2) the action condition (Ac-condition). According to (B3), belief requires understanding: to believe that *p*, one has to understand what *p*'s being the case amounts to. This condition may be termed the evidence condition (E-Condition).

In the example about Adam and Eve, the standard procedures for determining whether Eve is faithful support the conclusion that she is not: on the basis of the available information this is evident to everyone except Adam. The evidence condition makes one wonder whether Adam can "really" believe that his wife is faithful or whether he is just pretending. Nevertheless, Adam assents to the proposition that Eve has always been faithful to him, and is willing to defend this proposition against objections by constructing a protective barrier of auxiliary hypotheses around it. As to the second (action) component of belief, Adam may act as if Eve were faithful to him. Given Adam's interest in preserving his marriage, such action may in his case be (in most situations) optimal regardless of whether she is faithful or not. But this need not be the case: Adam's behavior towards Eve may change, even though he does not waver in his assent to the proposition that she is a good and faithful wife. Adam does not accept contradictory propositions, but he is not quite "unanimous" about Eve's faithfulness. The incoherence of Adam's beliefs is "hidden" at least in the sense that it does not involve the conscious assent to jointly inconsistent propositions.

Some of the examples which have been presented as evidence for the possibility of self-deception involving contradictory beliefs are based on the assumption that the presence of a belief can be detected by several indicators or criteria which can possibly conflict with each other. Mele reports Sackeim and Gur's experiment in which subjects denied that a tape-recorded voice was their own, but various physiological responses, for example, their GSR (galvanic skin response), indicated that they recognized the voice they heard.[35] The experimenters used the verbal report to determine that the subjects accepted a certain belief (viz., that the voice they heard was not their own) and regarded the behavioral indices as evidence that the subjects also held the contradictory belief.[36] The latter belief was thought to be "hidden," that is, the belief of which the subjects were unaware. The verbal criterion is the same as the assent condition (B1) above; the latter criterion is not the same as the action condition (B2) above, but a behavioral criterion of a different sort. In effect, Sackeim and Gur assume that the truth-values of belief sentences are determined by the following condition:

(BSG1) S believes that *p* if and only if some belief-indicator shows the presence of the belief that *p*.

If there are several logically independent indicators or criteria for the belief that P, it can of course happen that (BSG1) justifies contradictory belief ascriptions. But as Mele argues, this does not show that the subjects accept contradictory propositions.[37] We should conclude instead that (BSG1) is inconsistent with the construal of belief as a simple propositional attitude, expressible by "S believes that *p*" or "S believes that ~*p*." In reality, belief is a more complex phenomenon. From the standpoint of the propositional attitude theory of belief (the view that belief is an attitude towards a proposition), the phenomena of self-deception can be regarded as theoretical anomalies.

Notes

[1] This paper is an expanded version of comments delivered at the Florida Philosophical Association 2001 meeting.

[2] Jean Paul Sartre, *Being and Nothingness* (1943; New York: Philosophical Library, 1956) 49.

[3] Cf. Allen W. Wood, "Self-Deception and Bad Faith," *Perspectives on Self-Deception,* ed. B.P. McLaughlin and A. Oksenberg Rorty (Berkeley: U of California Press, 1988) 207.

[4] David Makinson, "The Paradox of the Preface," *Analysis* 25 (1965): 205-207.

[5] Bertrand Russell, *The Problems of Philosophy* (1912; London and New York: Oxford UP, 1946) 131; Frank P. Ramsey, "Knowledge," *Philosophical Papers*, by Ramsey, ed. D. H. Mellor (1929; Cambridge: Cambridge UP, 1990): 110-111.

[6] Henry E. Kyburg, *Probability and the Logic of Rational Belief* (Middletown, Conn.: Wesleyan UP, 1961) 70.

[7] H. Sackeim and R. Gur, "Self-Deception, Self-Confrontation, and Consciousness," *Consciousness and Self-Regulation: Advances in Research and Theory,* eds. G. Schwartz and D. Shapiro, vol. 2 (New York: Plenum Press, 1978) 150. Cf. Alfred Mele, *Self-Deception Unmasked* (Princeton and Oxford: Princeton UP, 2001) 81.

[8] Mele 2001. See also Alfred Mele, *Irrationality* (New York: Oxford UP, 1987).

[9] Moore's problem; see Roy A. Sorensen, *Blindspots* (Oxford: Clarendon Press, 1988) 1.

[10] Cf. G.E. Moore, *Ethics* (London: Williams & Nogate; New York: H. Holt, 1912) 125; G.E. Moore, "A Reply to My Critics," *The Philosophy of G.E. Moore,* ed. P.A. Schilpp (La Salle: Open Court, 1942) 540-543.

[11] Cf. Sorensen 55-56.

[12] "Rumsfeld: Bin Laden Hard to Catch," *The Miami Herald*, 26 Oct. 2001, final ed.: 1A.

[13] Rumsfeld quote from an (unidentified) article published in *USA Today* on 25 Oct. 2001, cited in "Rumsfeld: Bin Laden Hard to Catch," *The Miami Herald*, 26 Oct., 2001, final ed.: 1A.

[14] Mele 2001, 50-51.

[15] Cf. Mele 1987, 131-132.

[16] David Lewis, "Causation," *Journal of Philosophy* 70 (1973); David Lewis, "Postscripts to 'Causation'," *Philosophical Papers,* vol. 2, by Lewis (New York and Oxford: Oxford UP, 1986).

[17] Mele 2001, 122-123; Robert Audi, "Self-Deception vs. Self-Caused Deception: A Comment on Professor Mele," *Behavioral and Brain Sciences* 20 (1997): 104.

[18] Mele 2001, 51-52.

[19] Mele 2001, 123.

[20] Mele 2001, 123.

[21] Mele 2001, 123.

[22] David Pears, *Motivated Irrationality* (Oxford: Clarendon Press, 1984) 42.

[23] Mele 2001, 123.

[24] Mele 2001, 51.

[25] Imre Lakatos, "Falsification and the Methodology of Scientific Research Programmes," *Criticism and the Growth of Knowledge*, eds. I. Lakatos and A. Musgrave (Cambridge, U.K.: Cambridge UP, 1970) 132-138.

[26] Jonathan Evans, "Bias and Rationality," *Rationality: Psychological and Philosophical Perspectives,* eds. K.I. Manktelow and D.E. Over (London and New York: Routledge, 1993) 15; Jonathan Evans and David E. Over, *Rationality and Reasoning* (East Sussex: Psychology Press, 1996) 103.

[27] Mele 2001, 32.

[28] Cf. Evans; Evans and Over, ch. 5.

[29] B. Fischoff and R. Beyth-Marom, "Hypothesis Evaluation from a Bayesian Perspective," *Psychological Review* 90 (1983): 257.

[30] Evans and Over, 109.

[31] Peter Gärdenfors, *Knowledge in Flux: Modeling the Dynamics of Epistemic States* (Cambridge, Mass.: MIT Press, 1988) 17-18, 86-94.

[32] Cf. Evans; Evans and Over, 106.

[33] Y. Trope, B. Gervey and N. Liberman, "Wishful Thinking from a Pragmatic Hypothesis-Testing Perspective," *The Mythomanias: The Nature of Deception and Self-Deception,* ed. M. Myslobodsky (Mahwah, N.J.: Lawrence Erlbaum, 1997) 115; cf. Mele 2001, 29.

[34] Y. Trope and N. Liberman, "Social Hypothesis Testing: Cognitive and Motivational Mechanisms," *Social Psychology: Handbook of Basic Principles,* eds. E. Higgins and A. Kruglanski (New York: Guilford Press, 1996) 247.

[35] Mele 2001, 82.

[36] Mele 2001, 82; Sackeim and Gur, 173.

[37] Mele 2001, 82-83.

Works Cited

Audi, Robert. "Self-Deception vs. Self-Caused Deception: A Comment on Professor Mele." *Behavioral and Brain Sciences* 20 (1997): 104.

Dyke, Daniel. *The Mystery of Selfe-Deceiuing. Or A Discourse and Discouery of the Deceitfulnesse of Mans Heart.* London: Griffin and Mab, 1614.

Evans, Jonathan St. B. T. "Bias and Rationality." *Rationality: Psychological and Philosophical Perspectives.* Eds. K. I. Manktelow and D. E. Over, London and New York: Routledge, 1993. 6-30.

Evans, Jonathan St. B. T. and David E. Over. *Rationality and Reasoning.* East Sussex: Psychology Press, 1996.

Fischhoff, B. and R. Beyth-Marom. "Hypothesis Evaluation from a Bayesian Perspective." *Psychological Review* 90 (1983): 239-260.

Gärdenfors, Peter. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* Cambridge, Mass.: MIT Press, 1988.

Kyburg, Henry E. *Probability and the Logic of Rational Belief.* Middletown, Conn.: Wesleyan UP, 1961.

Lakatos, Imre. "Falsification and the Methodology of Scientific Research Programmes." *Criticism and the Growth of Knowledge.* Eds. I. Lakatos and A. Musgrave. Cambridge, U.K: Cambridge UP, 1970. 91-196.

Lewis, David. "Causation." *The Journal of Philosophy* 70 (1973): 556-67.

Lewis, David. "Postscripts to 'Causation'." *Philosophical Papers.* Vol. 2. By Lewis. New York and Oxford: Oxford UP, 1986. 172-213.

Makinson, David. "The Paradox of the Preface." *Analysis* 25 (1965): 205-207.

Mele, Alfred. *Irrationality.* New York: Oxford UP, 1987.

Mele, Alfred. *Self-Deception Unmasked*. Princeton and Oxford: Princeton UP, 2001.

Moore, G. E. *Ethics*. London: Williams & Nogate; New York: H. Holt, 1912.

Moore, G. E. "A Reply to My Critics." *The Philosophy of G. E. Moore*. Ed. P. A. Schilpp. La Salle: Open Court, 1942. 535-677.

Pears, David. *Motivated Irrationality*. Oxford: Clarendon Press, 1984.

Ramsey, Frank P. 1929. "Knowledge." *Philosophical Papers*. By Ramsey. Ed. D. H. Mellor. Cambridge: Cambridge UP, 1990. 110-111.

"Rumsfeld: Bin Laden Hard to Catch." *The Miami Herald*. 26 Oct. 2001, final ed.: 1A.

Russell, Bertrand. 1912. *The Problems of Philosophy*. Reset ed. London and New York: Oxford UP, 1946.

Sackeim, H. and R. Gur. "Self-Deception, Self-Confrontation, and Consciousness." *Consciousness and Self-Regulation*: *Advances in Research and Theory*. Vol. 2. Eds. G. Schwartz and D. Shapiro. New York: Plenum Press, 1978. 139-197.

Sartre, Jean Paul. 1943. *Being and Nothingness*. Trans. Hazel E. Barnes. New York: Philosophical Library, 1956.

Sorensen, Roy A. *Blindspots*. Oxford: Clarendon Press, 1988.

Trope, Y., Gervey, B. and N. Liberman. "Wishful Thinking from a Pragmatic Hypothesis-Testing Perspective." *The Mythomanias: The Nature of Deception and Self-Deception*. Ed. M. Myslobodsky. Mahwah, N.J.: Lawrence Erlbaum, 1997. 105-131.

Trope, Y. and A. Liberman. "Social Hypothesis Testing: Cognitive and Motivational Mechanisms." *Social Psychology: Handbook of Basic Principles*. Eds. E. Higgins and A. Kruglanski. New York: Guilford Press, 1996. 239-270.

Wood, Allen W. "Self-Deception and Bad Faith." *Perspectives on Self-Deception*. Eds. B. P. McLaughlin and A. Oksenberg Rorty. Berkeley: U of California Press. 1988, 207-227.

Reply to Commentators*

On *Self-Deception Unmasked* (Princeton and Oxford: Princeton UP, 2001)

Alfred R. Mele, *Florida State University*

I am grateful to my three friendly commentators for presentations that are bound to promote lively discussion. In the interest of leaving ample time for that, I will keep my reply brief. I will proceed in reverse order and start with Risto Hilpinen's comments. Incidentally, I agree with Risto that, in recent work on self-deception, relatively little attention has been paid to major historical literature on the topic, and it may be fruitful for people look more closely at this literature.

Risto asked about the confirmation bias, something I got mileage out of in the book. There is considerable evidence that it is a very common phenomenon. Here is a simple example. In one set of experiments, two different groups of people are asked to examine the same set of photographs of facial expressions. One group is asked to test the hypothesis, "Are these happy people?" The other group is asked to test the hypothesis, "Are these faces angry?" Most of the people asked the first question say "yes." Most of the people asked the second question say "yes." Why is that? It turns out that, much more often than not, people testing a hypothesis are much more sensitive to and receptive of confirming data than disconfirming data. In *Self-Deception Unmasked*, I review empirical evidence for the confirmation bias. I also argue that the bias can be triggered and sustained by desires—for example, the desire that one's son is not using drugs or that one's spouse is faithful—and that this helps to explain how it is that we sometimes believe what we would like to be true when we have stronger evidence that it is false.

If Risto's story about Adam and Eve is a story about self-deception, it describes an extreme instance of the phenomenon. Years ago, after I described a more typical case of self-deception about spousal infidelity, a student asked what sort of evidence of this kind of behavior would render self-deception about it impossible. My first thought was that catching one's spouse in the act would turn the trick, but I immediately started conjuring up a story in which even this might leave room for the belief that one's spouse is faithful and for self-deception about this. Later, I found a much better story—Isaac Bashevis Singer's "Gimpel the Fool." I summarize it in *Self-Deception Unmasked.* Here is a shorter summary.

One night, Gimpel, a gullible man, enters his house after work and sees "a man's form" next to his wife in bed. He immediately leaves—in order to avoid creating an uproar that would wake his child, or so he says. The next day, his wife, Elka, denies everything, implying that Gimpel was dreaming. Their rabbi orders Gimpel to move out of the house, and he obeys. In time, Gimpel begins to long for his wife and child. His longing apparently motivates the following reasoning: "Since she denies it is so, maybe I was only seeing things? Hallucinations do happen. You see a figure or a mannequin or something, but when you come up closer it's nothing, there's not a thing there. And if that's so, I'm doing her an injustice." Gimpel bursts out in tears. The next morning he tells his rabbi that he was wrong about Elka.

After nearly a year's deliberation, a council of rabbis allow Gimpel to return to his home. He is ecstatic, but wanting not to awaken his family, he walks in quietly after his evening's work. Predictably, he sees someone in bed with Elka, a certain young apprentice, and he accidentally awakens Elka. Pretending that nothing is amiss, Elka asks Gimpel why he has been allowed to visit and then sends him out to check on the goat, giving her lover a chance to escape. When Gimpel returns from the yard, he inquires about the absent lad. "What lad?" Elka asks. Gimpel explains, and Elka insists that he was hallucinating. Elka's brother then knocks Gimpel unconscious with a violent blow to the head. When Gimpel awakes in the morning, he confronts the apprentice, who stares at him in apparent amazement and advises him to seek a cure for his hallucinations.

Gimpel comes to believe that he has again been mistaken. He moves in with Elka and lives happily with her for twenty years, during which time she gives birth to many children. On her deathbed, Elka confesses that she has deceived Gimpel and that the children are not his. Gimpel the narrator reports: "If I had been clouted on the head with a piece of wood, it couldn't have bewildered me more." "Whose are they?" Gimpel asks, utterly confused. "I don't know," Elka replies. "There were a lot . . . but they're not yours." Gimpel sees the light.

This may be a case of self-deception. If it is, it is an extreme one. My point about it in the book is that even a case of self-deception as extreme as this does not require the machinery of a traditional conception of self deception—that is, simultaneously believing that *p* and believing that ~*p* and intentionally bringing it about that one acquires the belief one favors. Singer never tells us that while Gimpel believes that Elka has had an affair he also believes—at the same time—that she has not. Nor does he describe Gimpel as intentionally bringing it about that he believes in her fidelity. And there is no need to suppose that any of this is so in order to make sense of the story.

This leads me to Crystal Thorpe's commentary. Crystal's thesis is that I'm right about self-deception and my opponents are wrong, and she has an argument for this thesis that I don't advance. Her argument is that the traditional view of self-deception that I just mentioned makes self-deceived people seem much weirder than we take them to be.

I certainly won't disagree with this. The argument offers more support for my view. But what would happen if Crystal were to give this talk to an audience of Freudians and proponents of the traditional model of self-deception. They would say that self-deception requires believing that *p* while also believing ~*p*, intending to deceive oneself, and successfully executing that intention. And they have at least two options in responding to Crystal's argument. They can grant that, on their model, agents who deceive themselves are indeed weird and argue that the common-sense view of self-deceived people seriously underestimates their weirdness. Alternatively, they can argue that partitioned selves or whatever mechanisms they deem to be required for successfully executing intentions to deceive oneself and for simultaneously believing that *p* and believing that ~*p* really aren't so weird.

A promising response to Crystal's imagined critics, it seems to me, is to argue for the thesis that there is no need to appeal these mechanisms in explaining self-deception. And the best arguments for that thesis that I know of are mine. In any case, without an argument for this thesis Crystal runs the risk of begging the question against her opponents.

Finally, I turn to Peter Dalton's comments. In *Self-Deception Unmasked*, I offer a set of sufficient conditions for a person's entering self-deception in acquiring the belief that *p*. These conditions, as Peter indicates, are as follows:

1. The belief that *p* which S acquires is false.
2. S treats data relevant, or at least seemingly relevant, to the truth-value of *p* in a motivationally biased way.
3. This biased treatment is a nondeviant cause of S's acquiring the belief that *p*.
4. The body of data possessed by S at the time provides greater warrant for ~*p* than for *p*.

Peter suggests that a necessary condition of being self-deceived in acquiring the belief that *p* is that the person not believe that he reasoned incorrectly. Does Peter have the makings of a counterexample to my claim that my four conditions are sufficient for self-deception? If so, he should be able to produce a case in which my four conditions are satisfied, and even so, the person is not self-deceived because he doesn't satisfy Peter's condition.

Such a case would feature a person—call him Al—who *does* believe that he reasoned incorrectly. More precisely, Al believes that the reasoning on the basis of which he believes that *p* is incorrect. Now, for obvious reasons, Al's believing this would seem to make it hard for him to believe the conclusion of that line of reasoning—that is, *p*. There are two possibilities: (1) Al, who satisfies my four conditions, can believe that *p* even though he believes that the reasoning on which this belief is based is incorrect; (2) Al cannot believe that *p* in the circumstances at issue. Suppose that (1) is true. On that supposition, I challenge Peter to produce an instance of this possibility in which Al is *not* self-deceived in acquiring the belief that *p*! Given that in addition to satisfying my four conditions, Al believes that *p* despite believing that the reasoning on the basis of which he believes this is incorrect, we would seem to have a particularly perplexing, extreme case of self-deception on our hands and Peter would not have produced a counterexample. So suppose instead that (2) is true. Then Peter's necessary condition does not add anything substantive to my set of sufficient conditions. If (2) is true, no agent who satisfies my conditions fails to satisfy Peter's condition. That is, no agent who satisfies my conditions believes that the reasoning on which his belief that *p* is based is incorrect.

I'd like to make one last comment about my four conditions. I claim that being self-deceived in acquiring the belief that *p* requires that *p* be false. For me, then, the falsity of *p* is a necessary condition for such self-deception. Also, I deny that condition (4) is a necessary condition for self-deception. I argue that some cases of self-deception importantly involve a kind of blindness to evidence that is readily available. In some such cases, the evidence one actually possesses might favor *p*. Consider a zealous campaign worker for a presidential candidate. People might tell her that the candidate is corrupt because he's done x, y, and z. Instead of looking for documents that might give her evidence that he *has* done x, y and z, however, the campaign worker reads campaign literature in favor of her own candidate that takes a strong positive line on his moral character. Given further details, this may be a case of self-deception, even if the evidence the campaign worker possesses favors *p* over ~*p*.

*_This is an edited version of a transcription of an oral presentation to the FPA. I am grateful to the editors and their staff for their transcription._

Notes on Contributors

Sidney Axinn is Professor Emeritus at Temple University and Courtesy Professor of Philosophy at the University of South Florida. His areas of philosophical interest include Kant, social and political philosophy, military ethics, philosophy of science, and logic. His current work is on "sacrifice and value."

David Barnett is an undergraduate student at New College of Florida and winner of the 2000 Gerrit and Edith Schipper Undergraduate Award.

Peter Dalton is Associate Professor of Philosophy at Florida State University. His research interests include ethics, metaphysics, metaphilosophy, and the history of modern philosophy.

Risto Hilpinen is Professor of Philosophy at the University of Miami. He has published about 100 papers and edited several books in philosophical logic, epistemology, philosophy of science, and the philosophy of C. S. Peirce.

Jeremy Kirby is an instructor at Florida State University. His research interests include ancient philosophy, philosophy of science, epistemology, and metaphysics. He is currently working on a dissertation dealing with Aristotle's conception of dialectic and its relation to science.

Kirk Ludwig is Associate Professor of Philosophy at the University of Florida. Dr. Ludwig received his Ph.D. from the University of California at Berkeley. He works in the philosophy of mind, the philosophy of language, and epistemology. He was the President of the Florida Philosophical Association in 2001.

Alfred R. Mele, the William H. and Lucyle T. Werkmeister Professor of Philosophy at Florida State University, is the author of *Irrationality* (Oxford, 1987), *Springs of Action* (Oxford, 1992), *Autonomous Agents* (Oxford, 1995), *Self-Deception Unmasked* (Princeton, 2001), and *Motivation and Agency* (Oxford, forthcoming). He is the editor of *The Philosophy of Action* (Oxford, 1997) and coeditor of *Mental Causation* (Oxford, 1993) and *Handbook of Rationality* (Oxford, forthcoming). His primary research interests are in philosophy of action, philosophy of mind, and metaphysics.

Martin Schönfeld studied in Regensburg, Munchen, Georgia, and Indiana, where he earned his Ph.D. with F. Beiser and M. Friedman in 1995. Aside from *The Philosophy of the Young Kant* (Oxford

2000), he has published on the history of ideas, eighteenth century thought, environmental philosophy, and ethics. He is now writing "Kant's Development" for the *Stanford Encyclopedia of Philosophy* and translating (with J. Edwards) *Thoughts on the True Estimation of Living Forces* for the Cambridge edition of the *Works of Immanuel Kant*. He is currently Associate Professor of Philosophy at the University of South Florida and visiting professor at the Taipai Municipal Teachers College in Taiwan. He serves as the President of the Florida Philosophical Association for 2002.

Crystal Thorpe is Assistant Professor at the University of Florida. Her primary research interests include practical rationality, reasons for action, Kantian ethics and normative ethics.

Jennifer Uleman is Assistant Professor of Philosophy at the University of Miami. She works on moral and political theory and German Idealism. She is the author of several articles on Kant's practical theory, including most recently "External Freedom in Kant's *Rechtslehre*: Political, Metaphysical" forthcoming in *Philosophy and Phenomenological Research*. She is currently at work on a book about Kantian autonomy.

Byron Williston is Assistant Professor of Philosophy at Wilfrid Laurier University in Waterloo, Canada. He previously taught philosophy at the University of South Florida. Dr. Williston earned his Ph.D. at the University of Toronto. He has an edited volume on Descartes' Moral Philosophy coming out in fall 2002 with Humanity Books.