

## Epistemic Freedom<sup>1</sup>

---

Epistemic freedom is the freedom to affirm any one of several incompatible propositions without risk of being wrong.<sup>2</sup> We sometimes have this freedom, strange as it seems, and our having it sheds some light on the topic of free will and determinism.

I think that there are two equally important reasons why we seek an alternative to determinism as an account of how our actions come about. One reason is phenomenological: we just feel free. Determinism seems incompatible, in the first instance, with what it's like to be an agent. Our other reason for seeking an alternative to determinism is conceptual. We fear that if determinism is true, then we shall have no grounds for applying concepts such as responsibility and desert to ourselves and our fellows.

The conceptual reason for worrying about determinism has tended to take precedence in the writings of philosophers, but I think that the phenomenological reason deserves equal attention. My own view is that explaining what it's like to be an agent is just as interesting, philosophically, as finding room in the world for punishment and blame. And even those who disagree with me on this score should consider that the experience of freedom serves, in some philosophical theories, as a datum from which conceptual consequences are derived.<sup>3</sup> The conceptual problem of freedom thus becomes intertwined with the phenomenological problem.

<sup>1</sup> This chapter originally appeared in *Pacific Philosophical Quarterly* 70 (1989) 73–97 and is reprinted by permission of the University of Southern California. For comments on the material contained in this chapter, I am indebted to Paul Boghossian, Jaegwon Kim, Louis Loeb, Larry Powers, Dan Velleman, and Nicholas White. I also wish to thank various audiences to whom I have presented versions of the chapter, including the philosophy departments of Wayne State University, Indiana University, and The Ohio State University, as well as participants in a seminar on the philosophy of action that I conducted at the University of Michigan. Portions of this chapter appear in my *Practical Reflection* (Princeton: Princeton University Press, 1989), ch. 5.

<sup>2</sup> I use the vague word 'affirm' in this definition in order to encompass both assertions and beliefs. In this paper I shall confine my attention to epistemically free assertions, since epistemically free beliefs pose additional problems that I do not have the space to discuss. I allude to some of these problems in n. 17, below.

<sup>3</sup> See, e.g., C. A. Campbell, *On Selfhood and Goodhood* (London: Allen & Unwin, 1957), 158–79; Richard Taylor, *Metaphysics* (Englewood Cliffs, N.J.: Prentice-Hall, 1983), 23–50; and Roderick M. Chisholm, 'Human Freedom and the Self' in *Free Will*, ed. Gary Watson (Oxford: Oxford Univ. Press, 1982), 24–35.

This paper sketches a potential explanation for our feeling of freedom. It identifies a kind of freedom that we might have and that might cause us to feel free. The paper is therefore addressed primarily to the phenomenological problem of freedom, but it also has an indirect bearing on the conceptual problem. The freedom that I postulate is not causal but epistemic (in a sense that I shall define), and the result is that it is quite compatible with determinism. I therefore claim that insofar as we feel metaphysically free—free in a sense that would be incompatible with determinism—we are mistaking the epistemic freedom that we have for a kind of freedom that we may lack. This claim will lead me, at the end of the paper, to a projectivist account of moral responsibility. Ascriptions of moral responsibility, I shall suggest, should be treated in the same way as ascriptions of color or other secondary qualities.

The idea that our experience of freedom is occasioned by epistemic rather than causal freedom first appeared, I believe, in Hume's account of what he called "the *false sensation* . . . of the liberty of indifference."<sup>4</sup> According to Hume, we feel that our past doesn't determine our forthcoming behavior because of "a certain looseness which we feel in passing or not passing from the idea of one to that of the other." Here Hume seems to suggest that what makes us feel free is not a freedom to *do* any one of various things but rather a license to *think* any one of various things about what we are going to do. Hume contends that when we interpret this experience as a perception of gaps in causality, we're simply making a mistake.

Like Hume, I shall argue that our feeling of freedom is erroneous; and my diagnosis of the error, like Hume's, is that we mistake the license to affirm any one of various things about what we'll do for the possibility that we might do any one of those things. At the end of the paper, I shall discuss an occurrence of this confusion in the philosophical literature. First, however, I must establish that it is indeed a confusion; and my arguments to this end are somewhat different from Hume's.

### *The Openness of the Future*

The experience that I shall explain is often described, in particular, as the experience of openness in our future. Whenever we face a decision, we feel that our future is partly undetermined and thus leaves something for us to decide. This feeling seems to be a perception of real indeterminacy in the course of future events; how much indeterminacy is a difficult question. One might think

<sup>4</sup> David Hume, *A Treatise of Human Nature*, L. A. Selby-Bigge and P. H. Nidditch (eds.) (Oxford: Oxford Univ. Press, 1978), 408.

that our sense of deciding an aspect of the future intimates that there is no antecedent fact of the matter as to how that aspect will turn out. Alternatively, one might think that our sense of deciding an aspect of the future doesn't intimate that there is no fact about it but only that any such fact isn't causally determined by the present state of the world. Under either interpretation, the experience is taken to contain a denial of determinism, the first denial being considerably stronger than the second.

My thesis is that under either interpretation, the experience is an understandable illusion. Our sense of an open future is occasioned by a genuine indeterminacy, I believe, but the indeterminacy that occasions it is not the metaphysical indeterminacy that the experience represents to us. Our future is undetermined, I shall argue, in a way that explains our feeling of freedom without conflicting with determinism.

Of course, to say that indeterminacy doesn't conflict with determinism sounds like a stark contradiction. The reason why it isn't a contradiction is that talk of indeterminacy is ambiguous. There is a sense in which the indeterminacy of the future would falsify determinism, and then there is a sense in which it doesn't. These senses seem inseparable, but they can in fact come apart.

Here are the two different senses in which our future might be undetermined. On the one hand, there may be no particular way that the future is going to turn out—or at least, no way that's necessitated, under the laws of nature, by the present state of the world. In that case, the future would be metaphysically or causally open. On the other hand, there may be no particular way that we must describe the future as turning out, in order to describe it correctly—or at least, no way that's necessitated, under the laws of nature, by a correct description of the present state of the world. In that case, the future would be, as I put it, epistemically open. So formulated, these two kinds of indeterminacy seem inextricably linked. We naturally assume that if there is a way that the future will turn out, then that's the way we must describe it as turning out, in order to give a true description of it; and we assume that if the present will determine the future, under the laws of nature, then a true description of the present will determine, under the same laws, how we must describe the future in order to describe it correctly.

I shall argue that these assumptions, however plausible, are false. For in some cases, even if the future is going to turn out in a particular way, we don't have to describe it as turning out that way in order to describe it correctly, since there are several other, incompatible ways in which we would be equally correct to describe it as turning out. Similarly, even if the present determines how the future will be, a correct description of the present needn't dictate how we must describe the future as being in order to describe it correctly.

Here I have purposely made my claim sound implausible, to say the least. My reason for doing so is not that I wish to be credited with proving a startling claim. My reason is rather that I want to familiarize the reader with the strength of his own resistance to the idea of a divergence between the two indeterminacies that I have defined. My ultimate thesis, after all, is that we continually mistake epistemic indeterminacy, as I have defined it above, for causal indeterminacy; and this thesis gains support from any evidence that the two are generally perceived as inseparable. The reader's resistance to the thought of epistemic indeterminacy without causal indeterminacy is, if you will, a piece of psychological evidence for my claim that an experience of the one might be mistaken for an experience of the other.

In reality, the cases in which these indeterminacies diverge are not particularly startling, once they have been pointed out. Like most exceptions to specious generalizations, they hide in plain sight. Nevertheless, I shall continue to describe these cases in deliberately paradoxical-sounding ways, in order to heighten the reader's appreciation of their potential for causing confusion. I want the reader to realize that the confusion between causal and epistemic freedom, like the paradoxical sound of the cases in which they diverge, never entirely goes away, no matter how well it is understood.

### *Self-fulfilling Predictions*

Why would we sometimes be correct in describing the future in any one of several incompatible ways? The reason is that our descriptions would sometimes amount to self-fulfilling predictions.

Self-fulfilling predictions are more common than most people think. A famous example is a case discussed by G. E. M. Anscombe in the first pages of her book *Intention*. Anscombe imagines a doctor saying to a patient, in the presence of a nurse, "Nurse will now take you to the operating theater." The patient interprets this utterance as a straight-forward assertion of fact; the nurse interprets it as an implicit command and complies, thereby making the assertion true.<sup>5</sup>

<sup>5</sup> *Intention* (Ithaca: Cornell Univ. Press, 1963), 3. I shall follow Anscombe in saying that the doctor's utterance serves two functions, as both an assertion to the patient and a command to the nurse. However, nothing of significance hangs on characterizing this utterance as a command. Anscombe's story can be retold in such a way that the doctor's utterance is self-fulfilling even though its only illocutionary force is that of an assertion. For even if the doctor's utterance isn't intended or interpreted as a command, the nurse might still have motives for fulfilling it. The nurse might be afraid of undermining the patient's faith in the doctor, or of incurring the doctor's wrath by making him look like a fool. If the doctor is aware of these motives, he can count on his utterance to elicit the corresponding behavior whether or not it is an implicit command.

The point of the story, for my purposes, is that although the doctor is correct in asserting that the nurse will take the patient to the operating theater, he would have been equally correct in asserting that the nurse would take the patient to the lab, or to any other destination, within reason.<sup>6</sup> Insofar as the nurse stands ready to do whatever the doctor says, the doctor can truly assert any one of several incompatible things; and to that extent, he is epistemically free.

Of course, the epistemic freedom involved in this case is freedom enjoyed by the doctor in relation to the actions of the nurse. Because the subject of freedom here is not the agent, his freedom would seem to have little bearing on the question of free will. Yet the doctor in this case does, in a sense, decide what the nurse is to do; and his epistemic freedom attaches to the very utterance in which he formulates his decision. Perhaps, then, the present case is not as irrelevant to the question of free will as it seems.

I therefore plan to proceed as follows. First, I shall examine the epistemic freedom of the doctor in Anscombe's story, bringing to bear upon it various philosophical tools that have been developed for the analysis of predictability and determinism. I shall then suggest how the lessons learned in this case might be transferred to cases in which agents experience the distinctive feeling of freedom.

### *Epistemic Freedom and Determinism*

The doctor's epistemic freedom entails that the nurse's future is open from his perspective, in the sense that there is no one future that he has to predict in order to predict correctly. What's odd is that the nurse's future is open from the doctor's perspective even if it is predetermined. For if the nurse is predetermined to take the patient to the operating theater, in the case that I have described, then the reason must be that the doctor is predetermined to say so—which cannot change the fact that the nurse would take the patient elsewhere if the doctor said otherwise.

The doctor's epistemic freedom therefore amounts to the fact that he would be correct in predicting events that aren't actually going to occur. These events

<sup>6</sup> Within reason, I say. This qualification imposes a limit on the doctor's epistemic freedom—the limit of what the nurse would do if he predicted it. The nurse would probably take the patient to the lab, the lobby, the x-ray room, and several other places simply at the doctor's say-so. But if the doctor said "Nurse will now wheel you into the broom closet," he would be wrong. His prediction is therefore constrained by prior evidence about what a nurse, or this nurse, is likely to do on implicit instructions. I won't repeat this qualification, but it applies throughout the following discussion.

aren't going to occur only because he isn't going to predict them; and so if he did predict them, they would occur, and his predictions would be true.

Note, however, that although the doctor would be correct in predicting events that won't in fact occur, he doesn't directly confront the future nonoccurrence of the events in question. All that the doctor confronts is the present state of the world, which provides evidence about the events in question—potentially conclusive evidence, if determinism is true. What occasions the doctor's experience of freedom, then, is not that the future cannot determine how he should describe it but rather that the present cannot determine how he should describe the future. What makes him feel free, in short, is his freedom from the evidence.

The idea that an agent's descriptions of his own future actions are not constrained by evidence is hardly original with me. It has arisen periodically in the philosophy of action, especially since Anscombe's *Intention* and Stuart Hampshire's *Thought and Action*. Unfortunately, however, the philosophers who have discussed an agent's freedom from the evidence have usually linked it to epistemological or metaphysical views that are hard for the rest of us to accept. Some have contended, for example, that an agent's projections of his actions aren't constrained by evidence because they constitute a unique species of knowledge, "a kind . . . of knowledge to which the notion of evidence is irrelevant."<sup>7</sup> Others have contended that freedom from the evidence entails contracausal freedom of the will.<sup>8</sup> The idea of freedom from the evidence has therefore become associated, in the minds of most philosophers, with unacceptable conceptions of knowledge or causality. My hope is to make the idea respectable, by dissociating it from any suspect epistemology or metaphysics.

### *Freedom from the Evidence*

Anscombe says that the statement of the doctor in her story isn't "founded on evidence," even in its capacity as information to the patient. I believe that this characterization of the statement is mistaken. I say that the doctor's assertion to the patient is indeed based on evidence—namely, the evidence that the nurse is herewith getting implicit instructions to take the patient to the operating theater, that nurses usually understand and obey such instructions, and that this nurse has no inhibition against obeying this particular instruction. What

<sup>7</sup> Stuart Hampshire and H. L. A. Hart, 'Decision, Intention, and Certainty,' *Mind* 67 (1958) 1.

<sup>8</sup> Carl Ginet, 'Can the Will be Caused?', *Philosophical Review* 61 (1962) 49–55, and Richard Taylor, 'Deliberation and Foreknowledge,' *American Philosophical Quarterly* 1 (1964) 73–80.

Anscombe should have said, I claim, is that the evidence on which the doctor bases his assertion didn't and couldn't have dictated that assertion, since it wasn't at hand until the assertion was made. Until the doctor said that the nurse would take the patient to the operating theater, he had no particular evidence that the nurse would do so. All he had was evidence that the nurse would take the patient to the operating theater if he said so; and he had similar evidence about many alternative actions, each of which the nurse would have performed if he predicted it. Hence the doctor's evidence couldn't have dictated to him what to predict.

So stated, my argument sounds as if it relies on the doctor's ignorance, since his lack of compelling evidence sounds like something that would easily be remedied by more information. If determinism is true (as I shall henceforth assume for the sake of argument), then there is some action that the nurse is predetermined to perform; and which action the nurse will perform can in principle be extrapolated from the present state of the world. The nurse's future behavior is already in the cards, so to speak, and we assume that to anyone who is in a position to read those cards, they will dictate a determinate prediction. We are therefore inclined to think that if the doctor's prediction is underdetermined by his evidence, the reason must be that he can't see all of the cards—that his evidence is incomplete.

So we are inclined to think, but we're wrong. Let the nurse's behavior be causally predetermined; let the doctor know and fully appreciate all of the relevant laws and facts; let those laws and facts entail that the nurse will inevitably take the patient to the operating theater. Even so, the doctor is equally licensed to say "Nurse will take you to the lab" instead. Indeed, the evidence proving that the patient will be taken to the operating theater includes the very information that licenses the doctor to predict that the patient will be taken to the lab, since the way it proves that the patient will be taken to the operating theater is precisely by invoking the nurse's disposition to do what the doctor says—a disposition that would lead the nurse to take the patient to the lab if the doctor so predicted.

In the case of other observers, of course, the complete body of evidence indicates that they would be wrong to make an alternative prediction. It indicates that if they want to speak the truth, they must predict that the nurse will take the patient to the operating theater. The evidence thereby dictates a prediction to them, by showing that they mustn't diverge from it, on pain of error. But the evidence dictates this prediction to others by demonstrating that the nurse will fulfill it because of the doctor's making it. And to the doctor this evidence shows, not that he must make the prediction, on pain of error, but precisely the reverse—that no matter what he predicts, within reason, he can't go wrong. The evidence therefore contains one component that licenses the doctor to contradict what all of the evidence, taken together, conclusively

proves. This component of evidence shows that the doctor would be correct in predicting whatever he likes, within reason, irrespective of what the totality of evidence demonstrates is bound to occur.

The crucial component of the evidence is the part that supports various counterfactuals specifying the various places to which the nurse would take the patient if the doctor said so. This evidence includes facts about the nurse's dispositions to understand and obey implicit instructions, as well as facts about the absence of conditions inhibiting the nurse from taking the patient to the destinations in question. If the doctor is aware of this central component of the evidence, then he knows that although the complete body of evidence may indicate what the nurse will actually do, it cannot indicate that he must predict accordingly in order to predict correctly. The central evidence shows the doctor that even if the totality of evidence guarantees one outcome, he would still be correct in predicting others. It shows, in short, that he is epistemically free.

### *Causal Versus Epistemic Freedom*

Let me forestall a possible misunderstanding about what kind of freedom I am claiming for the doctor. I am claiming that one component of the evidence licenses the doctor to assert propositions even in the face of more extensive evidence guaranteeing their falsity; I am not claiming that the doctor's asserting one of those propositions is a physical or psychological possibility. The doctor has a license to say "Nurse will take you to the lab," but his saying it isn't causally possible given the antecedent facts.

After all, the doctor is in a particular conjunction of psychological states, which include his knowing that the nurse will do as he says, within reason, and his wanting the nurse to take the patient to the operating theater. Given the conjunction of these mental states, and the absence of any opposing forces, the doctor is predetermined to say "Nurse will take you to the operating theater." His license to say "Nurse will take you to the lab" is thus a license that he is predetermined not to invoke.

One might wonder about the point of having such a license. Indeed, one might wonder whether this license is anything more than a sham, issued only because the licensee is guaranteed never to invoke it. But the license involved here is not a sham. It is of course moot, in the sense that it isn't going to be invoked. But moot or not, it is still well worth having; it's worth having because it excuses the doctor from consulting evidence that would otherwise be pertinent to his prediction.

The evidence that the doctor is excused from consulting is the evidence that would indicate which destination, among the ones to which the patient might

reasonably be taken, is the one to which he will be taken in fact. For the purpose of making a true prediction, the doctor needn't investigate whether the nurse will take the patient to the lab rather than the operating theater, or to the x-ray room rather than the lab, so long as he is in possession of the central evidence indicating that the nurse will take the patient to whichever one of these destinations he names. To be sure, the doctor is predetermined to name the operating theater, and so his license to predict a trip to the lab is already moot. Yet to say that such a license is moot is not to say that it's a fake—that if the doctor attempted to invoke it, it would be revealed as invalid. No, the doctor's license to name the lab in his prediction is just as valid as his license to name the operating theater, since it indicates that naming the lab would result in an equally true prediction. The doctor is therefore entitled to name either destination without the support of evidence favoring it over the other.

One might think, alternatively, that the only reason the doctor needn't consult such evidence is that the prediction it would dictate is the one he's predetermined to make in any case. He can let his preferences determine his prediction, one might think, only because he knows that his preferences are guaranteed to yield that prediction which the evidence already supports.

One would be right that the doctor's preferences are guaranteed to yield such a prediction, but one would be wrong in thinking that this guarantee is of any significance. The fact is that the doctor needn't make a prediction that's congruent with the prior evidence about where the patient will be taken; for if he made a prediction contrary to that evidence, he would thereby have refuted it. The crucial evidence for saying that the patient will be taken to the operating theater is evidence that the doctor will say so. If he said otherwise, he would have proved this evidence false, thereby vindicating himself in contravening it.

As I have said, the doctor will not and cannot contravene this evidence. But to portray him as relying on the fact that he cannot contravene it is to imply that conformity to such evidence is a desideratum for him—which it isn't. Contravening the evidence about where the patient will be taken is impossible for him, but it's epistemically permissible; and precisely because it's permissible, he needn't rely on its impossibility.

#### *Complete versus Incomplete Evidence*

Of course, the doctor's license to make alternative predictions consists in what I have called the central component of evidence, the component showing that various predictions on the doctor's part would prompt various actions on the nurse's. If the doctor lacks this component of the evidence, then other evidence may well dictate a determinate prediction to him.

In order to illustrate this possibility, let me introduce a familiar prop of philosophical fiction. I shall ask you to imagine that there is a chronicle of the nurse's activities from birth to death, compiled and recorded in advance by a superhuman author. Imagine, further, that the doctor has obtained this book and has fully tested its reliability. For months, he has been surreptitiously trailing the nurse around the hospital in order to test the book's predictions, and he has found them to be exhaustive and unerring. Gradually, the doctor realizes that he is in possession of the genuine article, a book of life.<sup>9</sup>

One day the doctor enters a patient's room for a routine examination and finds himself face to face with the nurse whom he has been studying from a distance for all these months. During a lull in the subsequent conversation, the doctor extracts the nurse's book of life from his pocket in order to see what happens next. Leafing to the entry for the present date and time, he reads this: "Takes patient to operating theater." The doctor hasn't bothered to consider for himself what the nurse might do next; all he knows is that the book is always right. He thinks, "The operating theater, is it? Well, then, I'd better tell the patient"—whereupon he reports his discovery by saying, "Nurse will take you to the operating theater." The nurse, who knows nothing about books of life, interprets the doctor's utterance as a command and immediately complies.

In this version of the story, the doctor's prediction is dictated by the evidence available to him. In the absence of any other information, the book's proven reliability forbids him to gainsay the next entry. His information indicates that if he wants to speak the truth about the nurse's actions, he had better go by the book.

Yet the doctor's prediction is dictated by his evidence in this case only because that evidence is incomplete. In particular, the evidence dictating the doctor's prediction doesn't include information about his own role in causing the events predicted. The doctor therefore becomes the unwitting collaborator of the nurse's anonymous biographer, in somewhat the same way that Jocasta and Laius, the parents of Oedipus, became unwitting collaborators of Fate when they reacted to the prophecy about their infant son. The difference, of course, is that Oedipus' fate would have been fulfilled no matter what—or

<sup>9</sup> The classic discussion of books of life is in chapter 6 of Alvin I. Goldman's *A Theory of Human Action* (Princeton: Princeton Univ. Press, 1970). Goldman is primarily interested in the question whether the predictions in a book of life defeat themselves, if the subject reads them, by making him decide to prove them false. The same question is discussed in D. M. MacKay, 'On the Logical Indeterminacy of a Free Choice,' *Mind* 69 (1960) 31-40; David P. Gauthier, 'How Decisions are Caused,' *Journal of Philosophy* 64 (1967) 147-51; David P. Gauthier, 'How Decisions are Caused (But Not Predicted),' *Journal of Philosophy* 65 (1968) 170-1; and John O'Connor, 'How Decisions Are Predicted,' *Journal of Philosophy* 64 (1967) 429-30. I am interested in a different question, involving self-fulfilling rather than self-defeating predictions.

so the myth asks us to believe<sup>10</sup>—whereas, the biographer’s prediction about the nurse would not have been fulfilled without the doctor’s collaboration. And the doctor feels compelled by the biographer’s prediction only because he doesn’t realize that without his collaboration it won’t be fulfilled.<sup>11</sup>

For consider another version of the story, in which the nurse’s biographer has recorded, not just the nurse’s future actions, but also the chains of causes leading up to them—the chains down which the author himself must have peered in order to foresee the actions, in the first place. In this version, the doctor finds a more informative entry for the present date and time. He reads, “Hears the doctor say, ‘Nurse will take you to the operating theater,’ and is prompted to comply.” Reading this entry, the doctor realizes that the nurse will take the patient to the operating theater only because he’s going to say so. Does he now say to himself “I’d better tell the patient”? Does he feel that in order to warn the patient of impending events, he must echo the prediction written in the book? Surely not. He realizes that unless and until he repeats the book’s prediction, it won’t be fulfilled—and so he doesn’t have to repeat it in order to warn the patient. He also realizes that he can utter an alternative prediction without fear of a mistake.

Of course, even in this version of the story, the doctor says “Nurse will take you to the operating theater.” If he didn’t, the book of life would contain an error, whereas the story presupposes that the book is infallible. But when the doctor says “Nurse will take you to the operating theater” in this version of the story, he says it because he wants the nurse to take the patient to the operating theater, or for some such reason, and not because he thinks that he had better repeat what’s in the book in order to warn the patient of coming events.

Indeed, the nurse’s superhuman biographer must have realized that if he was going to reveal the causal inferences behind his entries in the book, then he couldn’t write an entry whose truth depended on the doctor’s thinking that he had better repeat it to the patient. The reason is that the doctor would never think this about an entry if it was explicitly based on the premise that he was going to think so. After all, the doctor wouldn’t think that such an entry was worth repeating for the patient’s information if he didn’t think that its stated premise was true; but its stated premise, in this case, would be that he’d think the entry was worth repeating for the patient’s information. Hence the doctor

<sup>10</sup> Consider how the myth would change if it were merely deterministic instead of fatalistic. In that case, the seer Tiresias would be the villain of the story, causing Oedipus’ downfall by uttering a self-fulfilling prophecy of it.

<sup>11</sup> Note that in this version of the story, the biographer’s prediction must have been based on his knowledge that the doctor would find the prediction compelling and consequently report it to the others, thereby unwittingly securing its truth. Evidently, then, the biographer could have written many other entries for the relevant date and time, in the knowledge that the doctor would feel compelled to report whatever was written, and that the nurse would be prompted to comply. Hence the biographer himself wasn’t compelled by prior evidence. But that’s another story.

wouldn't think that the entry was worth repeating unless he thought it was worth repeating; and so he wouldn't have to think so, if he didn't want to. If the nurse's book of life said to him, in effect, "The following prediction is true because you are going to find it worth repeating for the patient's information," he would read the prediction and rightly think, "Why should I repeat that? If I thought that the author had some *other* grounds for his prediction, I might think that it was true, and hence worth repeating to the patient. But if his only grounds for the prediction was that I would find it worth repeating to patient, then I don't."<sup>12</sup> The author must therefore have realized that in order to make a true prediction for which he could record his causal inferences, he would have to make a prediction that did not rely on the doctor's feeling compelled to repeat it.

What the author could rely on, and must have known he could rely on, was that if he alerted the doctor to his potential influence over the nurse, then the doctor would feel free, even in the face of a reliable prediction, to say what he wanted about the nurse's next action. Hence the author must have realized that so long as he made sure to write his entry in such a way as to reveal the doctor's potential influence, he could figure out what else to write simply by figuring out which action the doctor would want to prompt the nurse to perform.<sup>13</sup> In recording his causal inferences, the author must therefore have written, "Doctor wants nurse to take patient to the operating theater, and therefore says, 'Nurse will take you to the operating theater' . . ." and so on. Reading this, the

<sup>12</sup> Once again, the doctor might feel that the prediction was worth repeating if the evidence offered for it was incomplete. Suppose the doctor began to suspect that the nurse's biographer must have possessed, not only infallible foresight about his subject, but also hypnotic power over his readers. If the doctor then read "You are going to repeat, 'Nurse will take you to the operating theater,'" he might regard it as conveying a threat that the author was capable of carrying out. The doctor might then say to himself, "He's going to compel me to say that the nurse will take the patient to the operating theater. But if I say that, the nurse will do it. I'd better warn the patient." The doctor would thereupon blurt out, "Nurse will take you to the operating theater"—realizing too late that he had just been tricked into fulfilling the author's threat. Of course, if the author had recorded the psychological processes by which he expected to trick the doctor, the doctor would never have fallen for the trick.

<sup>13</sup> More precisely, the author would have needed to figure out, not just an action that the doctor would want to prompt, but an action that the doctor would want to prompt even after reading a prediction that he would do so. If the doctor is overwhelmingly averse to being anticipated, of course, there may be nothing that he would still want to do after reading a prediction of his doing it. If so, there may have been no prediction about the doctor that the author would have been correct to record, given that the doctor was going to read it.

My argument does not rely on the assumption that the author could always circumvent such perversity on the part of his subject. All I am claiming is that the author's recorded prediction, if explicitly based on the premise that the doctor would repeat it, could not put the doctor in the position of being epistemically compelled to repeat it—that is, of having to repeat it on pain of failing to make a correct prediction. In order to illustrate this claim, I have assumed that the author could have found some prediction that the doctor would want to repeat for reasons other than correctness, even if he had read a prediction of his making it. But this assumption is not crucial to my argument.

doctor thinks—not “I’d better tell the patient”—but rather “That *is* what I want to say. How clever of him!” And then he proceeds to say what he wants, and to say it only because he wants to.

Thus, the addition of further information to the first book of life undermines the book’s authority to dictate a prediction to the doctor, without undermining the book’s claim to be true. Once the book contains the central component of evidence, it shows that the doctor needn’t predict what it predicts in order to predict correctly, even though its prediction is, in fact, correct.

### *Questions and Answers*

*But suppose that the doctor reads the relevant entry in the more complete book of life on the previous evening. He is then compelled to conclude that when the time comes, he is going to say, “Nurse will take you to the operating theater,” and that the nurse is going to comply. The doctor therefore knows that the nurse is going to take the patient to the operating theater; and he can retain that knowledge until the moment of his utterance arrives. Surely, if he already knows that the nurse will take the patient to the operating theater, he isn’t entitled to say “Nurse will take you to the lab.”*

Surely he is. “Nurse will take you to the lab,” in his mouth, would still be just as true as “Nurse will take you to the operating theater,” and for precisely the same reasons. He is therefore fully entitled to say it.

*You mean that he’s entitled to say what he already knows to be false?*

Yes. The doctor is entitled to say what he knows to be false—not in the sense that he’s entitled to say something while knowing it to be false even as he speaks, but rather in the sense that something he now knows to be false is something that he’s nevertheless entitled to say, because it wouldn’t be false if he said it.

*But if it wouldn’t be false, then how could he have known it to be false?*

The answer is that if it wasn’t false, then he would not have known it to be false, after all. He does, in actuality, know it to be false, since he believes it to be false, with reliable justification, and it is false. But it’s false only because he isn’t going to say it. If he did say it, then it would be true, and his belief that it was false would not have constituted knowledge. In saying it, then, he would be saying something that he does (in actuality) know to be false but that he wouldn’t (in the circumstances) have known to be false. His utterance would be incompatible with what he actually knows but not with what he would know in that counterfactual case. Hence the doctor can know something to be false and still be entitled to say it simply because *if* he said it, he would not have known it to be false, after all.

There are two potential sources of confusion here. One is my claim that the doctor's asserting what he knows to be false would result in his not having known it to be false—a claim that seems to credit the doctor with the magical power of altering the past. All the claim actually attributes to the doctor, however, is the power of altering the epistemic status of past beliefs about events that still lie in the future; and this power requires no magic. If the doctor believes that the nurse won't take the patient to the lab, he will always have believed that proposition, even if he subsequently contradicts it. His having believed the proposition will henceforth be a fact about the past, which cannot be changed. But the doctor's having *known* the proposition will not yet be a fact about the past, since it involves a relation between the doctor's belief and the nurse's action, which is still to come. If the doctor were now to contradict his belief, his utterance would affect the nurse's behavior, thus affecting whether his earlier belief was true and hence whether it constituted knowledge. Thus, even though the doctor knows that the nurse won't take the patient to the lab, it can still be the case that if he contradicted that proposition, he wouldn't have known it, after all.<sup>14</sup>

Another potential source of confusion is that I seem to be saying that the evidence contained in the book supports mutually contradictory propositions—namely, the proposition that the doctor learns from the book and the alternative proposition that it licenses him to assert. But I am not saying that the book supports both propositions in the sense of showing both propositions to be true. Rather, I am saying that the book gives the doctor knowledge of one proposition by showing it to be true and licenses an assertion of the other proposition by showing that it would be true if asserted. Ordinarily, there is no difference between what evidence shows to be true and what it shows would be true if someone asserted it. Consequently, there is ordinarily no contradiction between the knowledge that evidence provides and the assertions that it licenses. In the case of self-fulfilling predictions, however, there can be a contradiction between what's actually true and what would be true if someone said it, since there are things that aren't true because the person isn't going to say them but that would be true if he did. And where there's a contradiction between what is true and what would be true if asserted, a person can know something on the basis of evidence that simultaneously licenses him to contradict it.

Here again, my point may seem academic, in the derogatory sense of the word, since it's about the potential truth of an assertion that the speaker is

<sup>14</sup> A more concise way of expressing the point of this paragraph is this. The claim that if the doctor contradicted what he knows, then he wouldn't have known it, sounds like an illicitly backtracking counterfactual. But it isn't one, because the antecedent, though phrased in the past tense, doesn't express a fact that's strictly about the past. (See David K. Lewis, 'Counterfactual Dependence and Time's Arrow,' *Noûs* 13 (1979) 455–76.)

predetermined not to make. But as I have argued, the potential truth of this utterance has implications for the speaker's actual position. Because the central evidence shows that he *would* be correct in making an alternative assertion, it shows that he *is*—in reality—entitled to make it. And if he is thus entitled to make any one of several different assertions, he may make one without consulting the evidence that would discriminate between them.

*You have conceded that even if the doctor told the patient "Nurse will take you to the lab," he would still have believed that the nurse would take the patient to the operating theater. You are therefore claiming that the doctor has a license to say something that he not only does believe to be false but also would have believed false even if he said it. Aren't you thereby claiming that he has a license to lie?*

The answer to this question has two parts. On the one hand, I haven't said that the doctor's asserting "Nurse will take you to the lab" wouldn't alter his belief: all I've said is that it couldn't alter the facts about what he antecedently believed. If the doctor said "Nurse will take you to the lab," and if he knew what he was saying, then he'd come to believe what he was saying, on the grounds that he was saying it. He would thus change his mind about what the nurse was going to do. Saying what he antecedently believed false would therefore entail changing his belief; and so it wouldn't entail saying something while still believing it to be false.

On the other hand, I recognize the possibility of the doctor's saying "Nurse will take you to the lab" without realizing that he was saying it, and hence without altering his antecedent belief. (The doctor might suffer a slip of the tongue and think that he was saying "operating theater" when he was actually saying "lab.") In that case, the doctor would indeed be saying something even as he believed that it was false. Yet even if we call the resulting assertion a lie, we must admit that it would be an inadvertent one. More importantly, we have to question whether the usual strictures against lying would still apply. Surely, the strictures against lying are based on the principle that a person ought to speak the truth, to the best of his ability. And in order to yield an injunction against saying something while believing it false, the principle that one should do one's best to speak the truth must be combined with the assumption that the best one can do, by way of speaking the truth, is to say what one justifiedly believes. In the case of self-fulfilling predictions, however, saying what one justifiedly believes is no more reliable a means of speaking the truth than saying what one justifiedly disbelieves. One can speak the truth by contradicting one's well-founded beliefs as well as by affirming them. Hence the principle that one ought to speak the truth, to the best of one's ability, doesn't yield an injunction against contradicting one's own beliefs.

*You have assumed that the causal inferences recorded in the nurse's book of life refer to psychological states and events. But if anyone actually wrote a book of life, he would probably base his predictions on physics, which yields better predictions than*

*psychology. And if the book cited the motions of particles and fields as causes of the nurse's actions, then the doctor might not realize that the causes cited were actually his own attitudes and utterances. Hence he wouldn't recognize his own role in causing the events predicted; and so he would once again be compelled to echo the book's prediction.*

I concede this point. But it's just another instance in which evidence would be compelling because it lacked the central component. In order to derive predictions about the nurse from premises about particles and fields, the author would have needed to know which microevents constituted which actions on the nurse's part. And if only he had cited similar correspondences for the doctor's actions, the doctor wouldn't have been in the dark about his causal role. The doctor's evidence would dictate his prediction, then, only because it failed to include the psychophysical relations necessary to complete the central component.

#### *Further Debate*

Whenever I present the thesis of epistemic freedom, I encounter opponents of two kinds. One opponent says that the thesis of epistemic freedom is obviously false; the other says that it is true, but trivially so. My first inclination is to answer either opponent by introducing him to the other. For how can the truth of a thesis be trivial, I wonder, if the thesis strikes some people as obviously false? And how can its falsity be obvious if it strikes others as trivially true? Of course, this reply never satisfies either party, and so the debate continues, along the following lines.

The opponent who calls the thesis of epistemic freedom trivial tends to compare my examples to cases of assertions that are more immediately and obviously self-fulfilling. In the eyes of this opponent, my claim that the doctor would be correct in saying "Nurse will take you to the lab," even though the nurse will actually take the patient to the operating theater, seems no more remarkable than the claim that even when the doctor isn't actually talking he would be correct in saying "I am talking." The opponent asks how such a trivial case can have interesting implications.

My initial response to this question is to admit, with pleasure, that the logic of "I am talking" is similar to that of "Nurse will take you to the operating theater" as it appears in Anscombe's story. The only significant difference between these assertions is that there is no obvious alternative to "I am talking" that the doctor would be equally correct in asserting, whereas there are clear alternatives to "Nurse will take you to the operating theater" that would be equally correct in the doctor's mouth.

Having admitted the similarity, however, I argue that the apparent triviality of the assertion “I am talking” does not prevent the case from having significant implications. The case demonstrates that a person can sometimes be licensed to make an assertion despite having conclusive evidence that it is false. For even if a person is not about to talk, and even if he has conclusive evidence to this effect, what deters him from saying “I am talking” cannot be this knowledge of the statement’s falsity, since that knowledge doesn’t imply that he would be wrong to make the statement. He may have evidence of conditions that will prevent him from saying “I am talking”—or from saying anything else—but that evidence cannot forbid him to say “I am talking.” A person can therefore feel entitled to say “I am talking” at any time without having consulted evidence about whether he is about to talk.

(At this point in the debate, I turn to my other opponent, who said that epistemic freedom was impossible, and I ask whether he can still think so, in light of this obvious case. Often he grants the existence of epistemic freedom and immediately joins forces with the opponent who calls it trivial.)

My first opponent now says that the implications I draw from the case of “I am talking” are well-known: epistemic freedom is therefore nothing new. Here I can only say that if it is nothing new, it is at least unfamiliar to some; and in any case, its relevance to the problem of free will has not been generally recognized. Its relevance to the problem of free will, I claim, is that when we have the distinctive experience of free will, we may be experiencing nothing more than epistemic freedom. Hence our feeling of freedom may be perfectly compatible with determinism.

The relevance of epistemic freedom to the problem of free will has not yet been demonstrated, however. I therefore turn to that task.

#### *Freedom in the First-Person*

From the perspective of the doctor in Anscombe’s story, the nurse’s forthcoming action is undetermined. I don’t mean that the doctor can truly say, “The nurse’s forthcoming action is undetermined.” If determinism is true, as I am assuming, then the nurse’s action is determined, and the doctor cannot truly say otherwise. What’s undetermined, for the doctor, is which action to say that the nurse will perform: it’s undetermined in the sense that there is no one action that the doctor must predict in order to predict correctly. Insofar as the action to be predicted by the doctor is undetermined, there is a recognizable sense in which the action itself is undetermined from his perspective. In the doctor’s eyes, the nurse’s immediate future is open, simply because he is entitled to predict that it will turn out however he likes.

I believe that the openness the doctor sees in the nurse's future is a perfect analog for the openness that he sees in his own. For I think that the openness that a person sees in his future is just the openness of epistemic freedom. Of course, the openness that the doctor sees in his own immediate future would ordinarily be thought of as reflecting his freedom to decide his next action rather than his freedom to predict it. But I believe that the traditional distinction between predicting and deciding breaks down in the case of self-fulfilling predictions.

Let me illustrate this phenomenon with a modest story of my own. My story is this. You go to a restaurant for lunch. The waiter gives you time to peruse the menu and then asks, "What will you have?" You reply, "I'll have a club sandwich." The end.

I take it to be uncontroversial that your utterance in this case expresses your decision about what to have for lunch. I propose, somewhat more controversially, that your utterance is also a self-fulfilling assertion. Just as the doctor asserted "Nurse will take you to the operating theater" in such a way as to bring about the patient's being taken there, so you assert "I'll have a club sandwich" in such a way as to bring about your having a club sandwich. You say that you'll have a club sandwich on the assumption that if you say so, then a club sandwich is what you'll get, and consequently what you'll have.

Some may object that your utterance in my story is a request or a command rather than an assertion; but this interpretation strikes me as inaccurate. If you said "I'll have a club sandwich" and the waiter replied "We're all out of turkey," then a natural thing for you to say would be "Then I guess I won't have a club sandwich"—which goes to show that you would regard your utterance as having misfired because of being false. (If you had started off with "Please bring me a club sandwich," the waiter's reply would not similarly lead you to say, "Then please don't bring me one.") I therefore feel safe in saying that your utterance purports to be true and, to that extent at least, qualifies as an assertion.

Now suppose that you were carrying a copy of your book of life, which you had tested and authenticated in the usual way. When the waiter asked "What will you have for lunch?" would you feel the need to consult the book before giving an answer that purported to be true? If you had already read that you were going to have a club sandwich, would you feel required to say so, on pain of speaking falsely? Would you be afraid to say "I'll have a chef's salad," lest your answer misfire? Certainly not. The reason, as I hope is now clear, is that you are epistemically free in relation to your reply, and you know it. When the question is "What will you have for lunch?" you are entitled to say whatever you like (within reason), because you'll have whatever you say.

*The Feeling of Freedom*

Now, my claim is that confronting the waiter's question "What will you have for lunch?" makes you feel that your future is open. You feel your future to be open, in respect to what you'll have for lunch, because you know that there isn't one, predetermined thing that you must say you are going to have, in order to speak the truth. You feel free to decide what you'll have for lunch because you know that there is, in your mouth, no unique true answer to the question "What will you have?" Yet as I have shown, the fact that there isn't one, predetermined thing that you must say you'll have, in order to speak the truth, is perfectly compatible with the fact that there is something that you're predetermined to have. There isn't a unique true answer for you to give, but there may still be a unique truth of the matter.

I do not claim, of course, that you are aware of the compatibility between your epistemic freedom and determinism; quite the reverse. The evident lack of a unique true answer for you to give in response to the waiter's question makes you feel that what you are going to have for lunch is metaphysically undetermined—that your luncheon selection is still open. But in feeling that your luncheon selection is open, you are mistaking epistemic for causal freedom. *All that's open is, not what you are going to have for lunch, but rather what you would be correct in saying you are going to have.* You mistake your license to say any one of various things about what you'll have for the possibility that you'll have any one of various things.

Here I am not making the familiar but, to my mind, less convincing claim that you feel free because of mere ignorance. According to that claim, the reason why you think that you might do any one of various things is simply that you don't know which of them you'll do. But the difference between not being sure what will happen and there being nothing that's sure to happen is perfectly clear in most cases. Why, then, should it elude you in this instance? My view is that if the experience of freedom is mistaken, it ought to rest on a more likely mistake. And although the difference between ignorance and metaphysical freedom is hard to miss, the difference between epistemic and metaphysical freedom is not. There being no unique answer to the question "What will I do?"—unlike your mere ignorance of the answer—is easy to mistake for there being no unique thing that you'll do.

*Can't the Will be Caused?*

This mistake is so easy to make, in fact, that it appears in the work of some philosophers, who have argued that a person's decisions require an openness that's incompatible with their being caused. One such argument is given by

Carl Ginet, as follows. A person cannot predict an action and then decide whether to perform it, according to Ginet, because his prediction would imply that the question to be decided was closed, and would thus preempt his decision. If a person's decisions were caused, however, there would be grounds on which he could predict them; and if he could predict the decisions, then he could also predict the resulting actions. But as Ginet has just claimed, a person cannot predict actions that he is to decide; and so Ginet concludes that a person's decisions must not be caused.<sup>15</sup>

As some critics have pointed out, the most that this argument can show is that in order for a set of conditions to be sufficient to cause a decision, they must be sufficient to prevent the decider from actually predicting it.<sup>16</sup> My criticism of the argument is different, however. I say that the argument fails to recognize the possibility of epistemic freedom.

Ginet's argument relies on the assumption that the agent's predicting an aspect of the future would preempt any decision about it, by closing the question how it would turn out. But I have argued that a person's prior knowledge does not close that question if he enjoys epistemic freedom. Even if you knew that you were going to say "I'll have a club sandwich"—and hence that you were going to have a club sandwich—you would still be in a position to say otherwise without risk of being wrong. And even if you then said "I'll have a club sandwich," as you knew you would, the reason would not be that your knowledge left you no permissible alternative. The evidence that you were going to have a club sandwich would have shown that you were going to have one because you were going to say so, and hence that you would have been equally correct in saying something else.

Ginet's argument would seem to imply, to the contrary, that if you knew that you were going to have a club sandwich, then you would be restricted to a single permissible answer when asked what you'd have. What I have shown is that even if you knew what you were going to have, you would be entitled to say that you'd have something else. From your perspective, then, the question "What will you have?" would remain open.

Naturally, this account of your feeling of freedom would be uninteresting—and, indeed, highly implausible—if it applied only to the small subset of

<sup>15</sup> Carl Ginet, 'Can the Will be Caused?'. Related arguments appear in Richard Taylor, 'Deliberation and Foreknowledge,' *American Philosophical Quarterly* 1 (1964) 73–80. See also John Canfield, 'Knowing About Future Decisions,' *Analysis* 22 (1962) 127–9, J. W. Roxbee Cox, 'Can I Know Beforehand What I am Going to Decide?,' *Philosophical Review* 72 (1963) 88–92, Andrew Oldenquist, 'Causes, Predictions and Decisions,' *Analysis* 24 (1964) 55–8, David Pears, 'Predicting and Deciding,' *Proceedings of the British Academy* 50 (1964) 193–227, David L. Perry, 'Prediction, Explanation and Freedom,' *Monist* 49 (1965) 234–47, Roy A. Sorensen, 'Uncaused Decisions and Pre-Decisional Blindspots,' *Philosophical Studies* 45 (1984) 51–6, and the works cited in n. 9, above.

<sup>16</sup> Roy A. Sorensen, 'Uncaused Decisions and Predecisional Blindspots.'

decisions that are formulated aloud as self-fulfilling assertions. But I believe that all decisions are self-fulfilling predictions of one sort or another, spoken or unspoken;<sup>17</sup> and so I believe that all decisions enjoy epistemic freedom. My purpose here is not to defend the proposition that decisions are self-fulfilling predictions, however.<sup>18</sup> My purpose has rather been to show that there is a kind of freedom that might explain the felt openness of your future without contradicting determinism. If your decisions were such as to enjoy this brand of freedom, then your experience of free will could be explained without resort to any metaphysical premises. Whether they actually do enjoy it is a question for another occasion.

### *Freedom as a Secondary Property*

Before concluding, however, I want to outline how my solution to the phenomenological problem of freedom may bear upon the conceptual problem. I

<sup>17</sup> In particular, I believe that most decisions consist in self-fulfilling beliefs. However, the epistemic freedom of beliefs is inextricably tangled up with issues that I wanted to avoid for the purposes of this paper. Let me say briefly what those issues are.

One important difference between utterances and beliefs is that our ability to say what we like is relatively uncontroversial, whereas our ability to believe what we like is not. (See Bernard Williams 'Deciding to Believe' in *Problems of the Self* (Cambridge: Cambridge Univ. Press, 1973); Stuart Hampshire, *Freedom of the Individual* (Princeton: Princeton Univ. Press, 1970); Brian O'Shaughnessy, *The Will: A Dual Aspect Theory* (Cambridge: Cambridge Univ. Press, 1980) vol. 1, 21 ff.) If we cannot believe what we like, then we have no use for a license to believe what we like about our future, which is what epistemic freedom would amount to in application to beliefs. I think, however, that the arguments against our ability to believe what we like are undermined in the case of potentially self-fulfilling beliefs. Most of those arguments involve the claim that a "belief" formed at will would not present itself to the subject as being reliably connected to the truth, and for that very reason would not qualify as a full-blooded belief. But a self-fulfilling belief can present itself simultaneously as having been formed at will and yet as reliably connected to the truth. Hence the main reason that has been adduced for denying that beliefs can be formed at will doesn't apply to those beliefs which would be epistemically free. This exception to arguments about believing at will is noted briefly by Jon Elster in *Ulysses and the Sirens; Studies in Rationality and Irrationality* (Cambridge: Cambridge Univ. Press, 1979), 48.

Another issue raised by the notion of epistemically free beliefs has to do with the ethics of belief. For as William James argues in 'The Will to Believe,' *Essays on Faith and Morals* (New York: New American Library, 1974), the possibility of self-fulfilling beliefs requires us to reconsider the commonsense notion that "It is wrong always, everywhere, and for every one, to believe anything upon insufficient evidence" [p. 39, quoting Clifford]. What needs reconsideration here, in my view, is the preposition 'upon', which turns out to be ambiguous. Talk about believing something *upon* evidence can refer either to believing the thing in response to evidence or to believing it with evidentiary support. Self-fulfilling beliefs, though subject to the requirement that they have evidentiary support, may well be exempted from the usual requirement of being prompted by evidence.

<sup>18</sup> I assert this proposition in 'Practical Reflection,' *Philosophical Review* 94 (1985) 33–61. I offer some defense in 'The Doxastic Theory of Intention,' in *Reasoning About Actions and Plans*, ed. Michael P. Georgeff and Amy L. Lansky (Los Altos, Cal.: Morgan Kaufmann, 1987). I defend the proposition at greater length in Velleman, *Practical Reflection*, ch. 4.

believe that my explanation of the phenomena yields a means of improving on the traditional compatibilist account of moral responsibility.

Traditional compatibilism about moral responsibility is unsatisfying, I think, because it insists on misconstruing what we mean when we say that a person has acted freely. Traditional compatibilists claim that moral responsibility can be said to require freedom, but only because 'freedom', in this context, means nothing metaphysical. When an ascription of responsibility is at stake, according to these compatibilists, the claim that a person has acted freely means, not that his action resolved some metaphysical indeterminacy, but only that it satisfied various conditions that are necessary for there being something to gain from subjecting the agent to praise or blame, reward or punishment. The problem with this theory is that it gets the meaning of the relevant discourse so obviously wrong. When we say that a person has acted freely, we mean that his action was metaphysically free, and no philosophical theory is going to convince us otherwise.

In my opinion, however, compatibilism can afford to leave the meaning of our discourse as it is. For if my explanation for the phenomena of freedom is correct, then metaphysical freedom is like a secondary quality, such as color. And in that case, compatibilism can take its cue from a projectivist account of secondary qualities, an account that interprets ascriptions of those qualities as systematically false and yet instrumentally justifiable.

The version of projectivism that I have in mind here is the one that says, not only that the color qualities that objects appear to have are merely projected onto them by our visual experience, but also that these qualities are the ones that are expressed by color predicates in ordinary discourse, and hence that our ascriptions of colors to objects are false.<sup>19</sup> According to this view, no external object has the quality that it is seen as having when it looks red or that it is said to have when it is called red.

Now, I believe that metaphysical freedom is a secondary property in the same sense. It's a property that we experience as being in the world, but only because we project it onto the world, by projecting a property of our predictions onto the actions predicted, thereby mistaking epistemic for metaphysical freedom. The metaphysical freedom that we consequently experience our actions as having, I want to say, is the property that we ascribe to those actions when we call them free. Hence our ascriptions of freedom, like our ascriptions of color, are systematically false: no action has the property that it is felt to have when it feels free or that it is said to have when it is called free.

I thus part company with the traditional compatibilists, who wish to interpret ascriptions of freedom in such a way as to make them come out true. I

<sup>19</sup> Paul Boghossian and I defend this version of projectivism in 'Color as a Secondary Quality,' in *Mind* 98 (1989) 81–103.

ally myself with the compatibilists, however, for the purpose of justifying—in my case, salvaging—the ascriptions of freedom that I have just condemned as false.

Here again, my view is modeled on the projectivist theory of color. The projectivism that I have in mind says that although color is an illusion, it is a sufficiently reliable and intersubjective illusion to be highly useful for everyday purposes. This illusion enables us to recognize objects, and to describe them in ways that make them recognizable to others, by adverting to the colors that the objects falsely appear to have. What's more, these purposes can be served equally well—indeed, better—if we ignore the distinction between an object's merely appearing to have a color (under standard conditions)<sup>20</sup> and its actually having one. We cannot go wrong, for the purpose of identifying objects, by believing the false testimony of our eyes or by reporting that testimony to others as if it were true, so long as everyone goes by the testimony of his eyes; for everyone's eyes conspire in precisely the same system of falsehoods.<sup>21</sup> Hence the falsity of color ascriptions is no reason for abandoning them or for assigning them a reformed meaning that would render them true.

I wish to vindicate ascriptions of freedom along parallel lines. I say that the feeling of metaphysical freedom, though illusory, does tend to track a real class of actions—namely, the actions in respect to which the agent is epistemically free. And I say that we have a legitimate interest in identifying actions belonging to that class, because they are the actions for which there will be something to gain from subjecting the agent to praise or blame, reward or punishment. The reason—which I cannot expound here—is that epistemically free actions are the ones that get performed only because the agent thinks he'll perform them, and wouldn't be performed if the agent thought otherwise; and such actions are precisely the ones whose performance could be influenced by the agent's application of the moral principles expressed in the practices of blame and punishment.

As I have said, I cannot defend this claim here. I state it only for the sake of showing where I think the compatibilist doctrine of responsibility can properly be applied. Its proper application is in explaining why we have a legitimate interest in the class of actions that falsely appear, from the agent's perspective, to be metaphysically free. Yet since everyone recognizes the actions in that class by their illusory metaphysical freedom—just as everyone recognizes objects by their illusory colors—there is no harm in treating the illusion as the truth for everyday purposes. Hence our ascriptions of freedom,

<sup>20</sup> Henceforth I omit this qualification.

<sup>21</sup> Here I am ignoring the possibility of spectrum inversion, of course. My reason for ignoring it is that it doesn't alter the value of describing objects by the colors that they appear to have. Indeed, the fact that it doesn't alter the value of such descriptions is precisely what makes spectrum inversion a problem, in the first place.

like our ascriptions of color, needn't be abandoned or reinterpreted. All that they require is an instrumentalist justification, along the lines suggested by compatibilism.

Of course, there are out-of-the-ordinary purposes for which these ascriptions, either of color or of freedom, can be disastrous. For example, if we attempt a scientific explanation of how our eyes detect the colors of objects, on the assumption that those colors are the properties that they appear to be, then we are bound to meet with failure.<sup>22</sup> If colors were as they appear, then our eyes would have to function as mere windows, transmitting those colors to some inner observer. This explanation is the one that informs our ordinary awareness of colors, I think, but it doesn't comport with the facts of physics or physiology.

Similarly, if we attempt a philosophical explanation of why blame and punishment are properly applied to free actions, on the assumption that the freedom of such actions is what it is felt to be, we are bound to meet with a corresponding failure. Again, we do have a commonsense notion of the connection between metaphysical freedom and desert—just as we have a commonsense notion of our eyes as windows on the colors of the world—but that notion won't stand up to philosophical scrutiny.

<sup>22</sup> Here I am accepting the projectivist account of visual experience. Other accounts attempt to argue that colors appear to be properties of a sort that external objects could have and that our eyes could detect. Arguments against such accounts appear in Boghossian and Velleman 'Color as a Secondary Quality' and 'Physicalist Theories of Color,' *Philosophical Review* 100 (1991) 67–106.