The Possibility of Freedom

John Maier

Draft of 8.03.08
Please do not cite without permission
Comments welcome: john.maier@anu.edu.au

People generally are so common in one's experience that it is natural to take them for granted, as presenting no puzzle or mystery, and to think only of such practical problems as arise in one's relationships to them, as fish must take other fish for granted, or as we take for granted the air around us and the stones at our feet . . . but some philosophic spirits, sometimes, are overwhelmed by a seeming discontinuity between themselves and the rest of physical nature, and they are sufficiently tormented by this apparent contrast to want to understand it and see what it implies.

– Richard Taylor, *Action and Purpose*

The Possibility of Freedom

Introduction

This essay is an extended discussion of a simple idea. The idea is that each of us has a varied though not unlimited range of actions that are *within his power*, or that are *options* for him: these are the actions that, as I will say, each of us is *free to* perform. The discussion attempts to make clear what exactly this idea is, and to what extent it can be justified.

This idea sometimes goes under the heading "free will," and we are sometimes said to "have free will" just in case we have a range of actions that we are free to perform. I avoid these terms because in ordinary and philosophical speech they normally refer to a diffuse set of concerns, only one of which I will address in this essay. It may be helpful to begin by briefly distinguishing some of these concerns and underscoring the aspect of freedom with which I will be concerned.

There are I think at least *two* sets of concerns which "free will" is used to denote, and so two ways in which the "problem of free will" can present itself. One way is by exposing an apparent disconnect between our practices and the justification for those practices, between what we *do* and what we *ought* to do. This problem arises most naturally in thinking about *moral responsibility*. Can it be fair to blame someone for a crime if he was determined, in some sense or other of "determined," to perform it? Can it be appropriate to *punish* him for what he is done? Since practices of blame and punishment (and of praise and reward) run through the life of social animals like ourselves, and since the threat posed by determinism is, whatever exactly determination comes to, a quite *general* one, the problem of free will, when developed in this way, seems to pose a radical threat to our ways of treating one another.

This development of the problem is a deep one, and it has received a correspondingly extensive measure of historical and contemporary attention. But there is another way of thinking about the problem of free will that is equally deep, and in some ways more ancient.[1] On this development, what is exposed is an apparent disconnect between our image of ourselves and the true facts about ourselves, between what *seems* to be and what *is*. In thinking of our pasts and futures, we think of various different actions that we *could have* taken or *can* take, and these thoughts inform our feelings about our pasts and our hopes and plans for our futures. But here too the apparent threat of determinism yields perplexing questions. If there is only one action that one really could have performed, does it make sense to regret what one has done? And if there is only one action that one really can perform, does it makes sense to deliberate about what one will do? Here too the problem is quite general. The threat posed by determinism does not merely threaten some particular aspect of our self-image. It rather challenges features of

[1] The Stoics, for example, seem to have been concerned in the first place with this problem, and only secondarily with questions of blame and praise. See Bobzien 1998.

ourselves so central that we do not know what it would be like to be an agent without them.[2]

When viewed in this second way, the problem of free will is most naturally associated with other problems that arise when we try to reconcile our prereflective picture of the world with the picture of it disclosed by the sciences. The former picture seems to include phenomenal properties, such as the redness of a particular apple, but there would seem to be no room for these in scientific picture. There are of course properties associated with colors, such as wavelengths, but *these* are not what we seem to see when we see that something is red.[3] The former picture seems to include evaluative properties, such as the wrongness of a particular action. But nothing in the natural world seems fit to play the role that we think of wrongness as playing.[4]

Similarly, I think, with freedom. Whether our world is deterministic or chancy, there seems to be no room in it for the properties that we think of ourselves as bearing. Like other objects, we are bearers of dispositions and propensities. But we also seem to stand in relations that other objects do not stand in. We have *options* – there are some actions that are *in our power* to perform, or that we are *free to* perform.[5] And these relations are nowhere to be found in the account of our world given by the sciences. It is *this* problem to which the present essay is addressed.

In this respect my concerns are somewhat different from much of the free will literature of the last several decades. This literature has been marked above all by a focus on the first problem, coupled with the thought that a satisfactory answer to it – a satisfactory account of how it can be fair for us to blame one another for what we do – need not require that agents be genuinely free to act otherwise than they actually do. Cases involving certain sorts of overdetermination, coupled with more general concerns about the demands of the reactive attitudes, have tended to support the thought that agents may deserve blame and perhaps also praise for what they do *even if* they are never free to do otherwise. The claim that agents free to do otherwise has therefore come to be seen as in at least one way less significant.[6]

I think that this literature leaves what I have called the second problem – the problem of reconciling our sense that we are free with the picture of the world disclosed by the sciences – very much open. For our grounds for thinking that we are free to do otherwise are, on my view, quite independent of concerns about moral responsibility. They arise,

---

[2] Compare these two ways of developing the problem with David Velleman's distinction between the "conceptual" and the "phenomenological" problems of freedom in Velleman 1989.

[3] Moore 1903.

[4] Mackie 1977. While in Moore these sorts of anti-reductive concerns lead to Moore's distinctive realism, in Mackie they are taken to support the error theory. That these sorts of concerns lead to this sort of whipsaw between reification and elimination is I think a quite general phenomenon, though in the case of freedom the tendency has been towards elimination.

[5] For the distinction between this sort of relation and the propensities of objects, see Taylor 1960 and Hampshire 1975.

[6] I think of this trend beginning with Strawson 1956 and Frankfurt 1969. More recent developments include Wolf 1990, Wallace 1994, and Fischer and Ravizza 1998.

rather, from a quite fundamental aspect of the experience of being an agent. So I will have little to say in what follows directly about moral responsibility, but I will argue that – *whatever* the true view about moral responsibility – we have grounds for thinking that we are free to perform actions that we do not perform.

I will also argue that our being free in this respect is compatible with the truth of determinism – in at least one sense of "compatible." In this way, the view that I will develop here satisfies the aspirations of traditional compatibilism.[7] In other ways my defense of compatibilism is less than traditional. One way, which will become clear presently, is that I reject any attempt to give a *reduction* of freedom into more basic terms. Another way will emerge towards the end of the essay. This will turn on distinguishing the epistemic aspirations of traditional compatibilism from its metaphysical aspirations. Only the former, I argue, are ones that we can reasonably hope to achieve.

Before coming to all of that, I want to make as clear as possible what I am talking about when I talk about "freedom." This is what I will do in the next chapter. Accusations that discussions of free will are hopelessly obscure are perhaps as old as these discussions themselves. In many cases, I think, these accusations are just. Nonetheless, I think there is at least one set of problems that goes under this heading that can be framed in a way that satisfies any reasonable criterion of clarity. On my view, this is a question about the metaphysics and epistemology of a *relation* – the freedom relation – with which we are all already acquainted. In the next chapter I will try to make explicit what exactly this relation is.

[7] Though it is clear enough what traditional compatibilism is, it is less clear whom we should count as the traditional compatibilists: Hobbes, Locke, and Hume of course bear mention here, but their actual views on this issue are nuanced in ways that make it difficult to assign a particular thesis to them. Perhaps the clearest statement of the view which I will eventually defend occurs in the opening paragraphs of Lewis 1981.

The Possibility of Freedom

Chapter 1: The Freedom Relation

1.1 The concept of freedom

In discussions of freedom, the demand is often made to say *in what sense* one is using the phrase "freedom." In one respect, this demand is a reasonable one. This term is used in different ways by different philosophers, and discussions that do not get as clear as possible about such phrases risk devolving into merely verbal disputes. Indeed, the main purpose of this chapter is to say as clearly as possible what *I* mean when I speak of "freedom."

Nonetheless, there is one respect in which this demand can read too strongly. Read in this stronger way, the demand is not merely to explain how one is using a certain term, but to offer one's *conception* of freedom. On this view, there are a variety of conceptions of freedom, some more demanding and some less, and the task of the philosopher is to defend the claim that his conception of freedom is in some respect or other better than those of his competitors. Read in this way, I think the demand is misplaced. I will not offer any particular conception of freedom and argue that it is somehow preferable to other conceptions. When I speak of freedom, I take myself to be referring to something that is already implicit in our speaking and thinking about agency.[8] I do not take myself to be *introducing* the idea of freedom, though I will try to speak more precisely about it than we normally do. In doing this, however, I take myself to be fixing on something that is there *already*. I am simply trying to pick out the *concept* of freedom.[9]

1.2 Acting freely and freedom to act

There are several different questions one might be asking when one asks whether or not we are free with respect to our actions. One question is a question about the actions we in fact perform. One might ask, of these actions, whether at least some of them are ones that we perform *freely*. The other question is about the actions that we do *not* perform. One might ask, of these actions, whether at least some of them are ones that we are *free to* perform. My concern in this essay will be with the *second* of these questions: are we *free to* perform actions that we do not perform? It is however worth dwelling briefly on the first of these questions, in order to distinguish it properly from the question that I *will* be answering as well as to flag several issues that would be treated in a fully adequate

---

[8] And not merely something to which our ordinary ways of thinking and speaking *approximate*. This is one of the more significant differences between the present approach and some strands of traditional compatibilism, which sound what we might call an *ersatzist* theme. Thus Jackson 1998 writes: "What compatibilist arguments show, or so it seems to me, is not that free action as understood by the folk is compatible with determinism, but that free action on a conception near enough to the folk's to be regarded as a natural extension of it . . . is compatible with determinism" (pp. 44-45). Whatever the merits of this sort of ersatzist program, it is not mine. I will eventually argue that freedom *simpliciter* is compatible with determinism. (This need not entail that all of the folk's beliefs about freedom are correct ones; indeed, the idea that freedom conflicts with determinism may *itself* be a false belief of the folk).

[9] On the distinction between concepts and conceptions, see Rawls 1971.

account of free action but will remain largely unaddressed here. What might someone mean in asking whether, of an action that someone does, whether it is one that he does *freely*? Let us call this the *adverbial question*. There are I think several things that might be meant by the adverbial question.

In asking the adverbial question, we might be asking simply whether someone is *free to* do otherwise. Thus it is sometimes said that a free action is one which it was in someone's power *not* to perform. This is the sense in which David Lewis is using the term "free act" when he writes: "I have just put my hand down on my desk. That, let me claim, was a free but predetermined act. I was able to act otherwise, for instance to raise my hand."[10] In this sense, there is no divergence between the adverbial question and the question that I am interested in. The adverbial question – do we ever act freely? – is just an alternative way of stating my question – are we ever free to act otherwise? There are however several readings of the adverbial question on which it *does* come apart from the question that I am interested in. Let me consider three such readings.

First, in asking whether an action was performed freely, we might be asking whether something that appears to be an action really is an *action*. This is the sort of thing that we mean when we ask, of someone whose bodily movements are sometimes caused by electric currents administered by a doctor, whether he moved his arm freely in some particular case. In this sort of case, I think, our question is merely whether his arm-raising was really an action or whether it was merely an effect of the doctor's intervention. The question of what it takes to act freely in this first sense is addressed by theories that attempt to give an account of the distinction between actions and mere events. I shall not be offering a theory of *this* sort in this essay. I will take the distinction between actions and mere events to be one of which we can, somehow, give an account.

Second, what might be meant is whether, given an agent has indeed performed some action, whether that action meets some further, more demanding, standard. This is the sort of thing we mean when we ask, of a heroin addict, whether his purchase of heroin was done *freely*. In this sort of case, I think, our question is whether this action one with which he "endorses," or with which he "identifies," or one which flows from his "true self."[11] These are elusive but intuitively compelling questions, and I take it that one of the main achievements of work on free will in recent decades has been to commend them to our attention. Nonetheless, this sort of more demanding standard is not one of which I will try to give an account in this essay, and though there are doubtless such distinctions to be made among actions, nothing I will have to say here will depend on drawing any such distinction.

Third, what might be meant is whether, given the agent has performed some action, whether it is one he is *morally responsible* for. This is the sort of thing we mean when we ask, of some bank robber, whether he held up the bank freely – or whether, for example, he did so under some coercive threat to which he reasonably acceded. In this sort of question, we are asking whether the action in question satisfies the conditions of

[10] Lewis 1981.
[11] See, respectively, Frankfurt 1971, Frankfurt 1987, and Wolf 1990.

moral responsibility, whatever exactly those might be. As I have already said, these sorts of questions will not be foregrounded here. I will offer no theory of the conditions of moral responsibility, and questions about moral responsibility will arise only insofar as they bear on the question with which I am concerned.

That question is the latter question above: whether I am *free to* perform actions that I do not perform. This is the sort of thing we mean when we ask, of someone who has just raised his right arm, whether he was free to raise his left arm instead – or whether, for example, his left arm is paralyzed or bound to his desk. For my purposes, it is necessary and sufficient this his raising of his left arm would be an action, whether it would be something that he *did* – that it would not be a mere bodily movement, in the way in which the reflex movements caused by electrical currents are. That is, all that matters is whether his action would be a free one on the first reading of the adverbial question.

Of course, it may be that these nominal distinctions *in fact* coincide. For example, it may be that to be an action *just is* to be an event that someone is morally responsible for, in which case the first and the third reading of the adverbial question would have the same answer. And each of these readings of the adverbial question may in fact coincide with the question that I am interested in. For example, it could be that to be morally responsible for an action *just is* to be free to not perform that action. But each of these is a substantive claim that would have to be evaluated on its merits. We should begin with the thought that there are at least four separate questions one might ask about free action: the question that I am interested in, namely whether someone is free to do otherwise, and what I am calling the three readings of the adverbial question. Any claim that these distinctions in fact coincide is a *theory* about the nature of free action, not merely a matter of meaning.

1.3 The freedom relation

The previous section explains some of the questions that I will *not* address in this essay. This section explains the question that I will address.

What do I mean by "freedom"? I conceive of the topic of inquiry here not as a certain way of *talking* about agents, nor of *thinking* of agents. The topic of inquiry here is a *relation*, one which is often successfully picked out by our ways of speaking and thinking about agency, but which does not *depend on* the existence of those ways of speaking and thinking. In this sense, this essay is in the first place an exercise not in linguistics or psychology but in metaphysics.

What relation do I have in mind? I think the best way of latching onto this relation is by way of *example*. Consider now your right arm. I think that, if you have a normally functioning mind and body, at least two claims are true of you. You are, provided your arm is not tied down or otherwise restricted, now *free to* move your arm upwards. And you are, provided your arm is not atop a desk or otherwise impeded, now *free to* move your arm downwards. *This* relation, the one that you now stand in to both moving your arm upwards and moving your arm downwards, is what I am calling the freedom relation.

It will be objected that there are a variety of distinct relations that one stands in to these actions – which of them is the freedom relation? The best answer to this worry is simply to iterate the method of examples: for any such relation, there will be a case where it comes apart from the freedom relation, and we can distinguish it from the freedom relation by considering such a case.[12] Consider the relation, for example, of *believing that one is free to* do something. This is a relation that you *also* stand in to moving your right arm upwards and to moving your right arm downwards. To distinguish this relation from the freedom relation, we need only consider cases where you have false beliefs. Consider the case where your arm is, unbeknownst to you, bound to the floor by finely woven strings. In such a case, you *believe you are free to* raise your arm, but you are not *free to* raise your arm.

It would be nice to have a more formal way of specifying the relation that I am interested in. But I think that this method – the deployment and, if necessary, redeployment of the method of cases – is both the best method available in this case and perfectly adequate as a way of picking out the freedom relation. So this is my answer to the demand posed at the beginning of section 1.1, to say in what sense I am using the term "freedom." I am using it to pick out a relation which we can identify through reflection on particular cases.

1.4 More on the freedom relation

The freedom relation is a relation between agents and types of actions that they do not perform. Its canonical form is:

(FR) S is free to A

Where 'S' picks out an agent and 'A' a type of action that S does not, in fact perform. In this section I will say something more about the relata of this relation.[13]

I take the domain of 'S' to include agents: the sort of beings who act.[14] I shall not have much to say about the domain of 'S', beyond noting that it includes you and me as canonical instances.

---

[12] There would be no such case for some relation R if that relation were necessarily coextensive with the freedom relation: if for any pair of agents and act-types, an agent stands in R to that act-type just in case he is free to perform it. In Chapter 2, I will argue that there is no such relation other than the freedom relation itself.

[13] There are some further issues about specifying the freedom relation that I do not address here. One concerns time. On the plausible assumption that an agent can be free to perform some action at one time, but not free to perform that same action at some other time, (FR) would seem to require a restriction to times. The most straightforward way to do this is by making (FR) a three-term relation, holding between an agent, a place, and a time. Alternately, we might say that our plausible assumption is not correct, and that strictly it an agent-at-a-time who is free to perform an action, or that the proper specification of action-types requires a reference to times. In what follows I speak as if it is the third of these alternatives that is correct, though for present purposes we might just as well take one of the first two alternatives instead.

[14] It may be that not every being who is capable of acting stands in the freedom relation to some action. Perhaps the demands for freedom are set higher than the demands for (mere) action.

I have a bit more to say about the domain of 'A'. I said above that its domain is types of actions which an agent does not perform. There are two questions to be asked about this formulation. First, why *types* of actions? Second, why types of actions an agent does *not* perform? Let me take these questions in turn.

To say that freedom is a relation to types of actions is to say that it is not a relation to any *particular* action but rather to a *class* of actions, a class which can typically be specified by an action description. In this respect the freedom relation is different from the *action* relation. To act is, on one standard picture, for a relation – the relation of being caused in the right way by one's desires and beliefs – to hold between an agent and some particular event in the world.[15] But there *is* no event in the case of the freedom relation. This is why I say that freedom, unlike action, relates agents and *types* of actions. This types may be as finely specified as one likes, so one may be free to raise one's arm, but not free to raise one's arm at a certain velocity, or within a certain span of time.

Why the restriction to actions one does *not* perform? It is clear that the domain of this relation seems to include *some* actions that one does not perform: for example, the action of now raising one's arm upwards mentioned earlier. But why should it include *only* actions one does not perform? Couldn't we just as easily say that the freedom relation is a relation to actions one does not perform *and* to actions that one does perform – where one is free to perform a type of action if one performs some particular action that satisfies that type?

I think this would be a mistake. There is of course a relation that we can construct from the freedom relation conjoined with the action relation, but this relation is a gerrymandered one. One way of arguing for this point is by considering a debate that I will keep to one side for most of this essay: the debate between the *compatibilist* and the *incompatibilist* about freedom.[16] Whatever exactly this debate comes to, it is a debate about the actions an agent does *not* perform. Our present formulation allows us to capture this aspect of their debate: the incompatibilist denies, in some sense, that the freedom relation would ever be instantiated in a deterministic world. So the compatibilist and the incompatibilist are disagreeing about the demands of the freedom relation. And this seems correct. Yet on the more permissive reading of the freedom relation, which includes actions that an agent does perform, the incompatibilist does not deny that this relation is sometimes instantiated in deterministic worlds – provided he grants, as at least some incompatibilists do, that determinism *per se* is no obstacle to action. The present formulation thus captures the intuitively correct idea that the freedom relation, whatever else it is, is that about which the compatibilist and the incompatibilist disagree.

This point may seem to be merely formal: why not just grant that there are two relations here, and that it is merely a matter of convenience which we call "the freedom relation"? This too, I think, would be a mistake. The failure to recognize that the freedom relation is essentially a relation to types of action that one does not perform can lead to a confused

[15] The picture, though not all of the terminology, is that of Davidson 1963.
[16] I will turn to this debate in Chapter 5.

account of the *epistemology* of the freedom relation. So I will argue, at least, in what follows.[17]

1.5 The ability relation

The freedom relation is typically picked out in English by phrases other than "is free to." These include the phrases "is able to" and "has the ability to" as well as by the troublesome modal auxiliary "can." The proper semantics for this part of English are far from clear, and I will not try to make any substantive progress on this front here. The main task of this section and the next are to *forestall confusion* by spelling out carefully the way that the freedom relation is referred to in English. In the next section I will highlight some varieties of imprecision that enter into our talk about freedom. In this section I want to distinguish the freedom relation from another relation which is often picked out by the same phrases, but which is distinct from the freedom relation and whose metaphysics raise a different set of questions. This is what I will call the *ability relation*.

The distinction between the freedom relation and the ability relation is, like the freedom relation itself, best made clear by way of examples. Consider some agent who was brought up in France as a native speaker of French.[18] Such an agent has the ability to speak French if anyone does. But if such a speaker is gagged, then intuitively he is not free to speak French. Speaking French is not, under the circumstances, something he can do. (This is *why* we put gags on people). Such an agent is a paradigm example of someone who stands in the ability relation to speaking French but does not stand in the freedom relation to speaking French.

No analysis of the ability relation will be forthcoming here. Nonetheless, let me make a couple of remarks about it. Let us take the canonical form of the ability relation to be:

(AR) S has the ability to A

Where the domain of (AR), like the domain of (FR), is agents and types of actions they do not perform, respectively. It seems that any adequate analysis of the ability relation would have to respect two plausible constraints:

(i) It is possible that S has the ability to A but S is not free to A
(ii) If S has the ability to A, then this is so in virtue of the intrinsic properties of A

(i) is just a more general restatement of the point made above about a gagged French speaker. (ii) I think is an intuitively compelling constraint on ability ascriptions, one which parallels a similar constraint on dispositions ascriptions.[19] When one has an ability

---

[17] Especially in Chapter 3.

[18] Peter van Inwagen uses this example to make the same sort of distinction that I am making in van Inwagen 1983.

[19] For a useful discussion of the intrinsicality constraint on dispositions, coupled by the way in which it allows for dispositions that are in some sense "extrinsic," see Fara 2001, chapter 5. One subtlety not

this is so because of how one intrinsically is, not because of the circumstances in which one happens to find oneself.

If (ii) is true, then it points to a deeper disanalogy between the ability relation and the freedom relation. For the freedom relation does not seem to be subject to this sort of intrinsicality constraint. If that is correct, then it allows us to construct another sort of example where ability and freedom come apart. The example of the gagged French speaker showed that standing in the ability relation to an action is *not sufficient* for standing in the freedom relation to that action. But the following example seems to show that standing in the ability relation to an action is *not necessary* for standing in the freedom relation to that action.

Consider two ways in which a benevolent genie might make an agent free to perform some skilled action, say hitting a serve in tennis. One way is by granting him, temporarily, the physical and mental capabilities required to hit a serve. Another way is by keeping his physical and mental capabilities as they are, but intervening in his environment so that whatever actions he undertakes will result in his hitting a serve – for example, by altering the trajectory and velocity of the tennis ball so that, when he strikes it, it will result in a successful serve. In the former case, I think, the benevolent genie makes the agent free to hit a serve by bestowing upon him the ability to hit a serve. In the latter case this is not so. The agent does not have the ability to hit a serve, for he fails to satisfy condition (ii). Nonetheless, the genie's interventions put him in a situation such that he is free to hit a serve – in virtue of the genie's assistance, he is, in these circumstances, free to hit a serve. So standing in the ability relation to an action does not even appear to be necessary for standing in the freedom relation to that action.

Examples like this one will recur in the next chapter, where I will argue that the freedom relation cannot be reduced to the conjunction of the ability relation with some other condition. (It should be clear that this simple reduction is already refuted by the point just raised, namely that ability is not even necessary for freedom). Let me close by noting that the distinction between the freedom relation and the ability relation is not at all novel. This distinction is labeled, however, with a depressingly non-uniform set of terms. The most standard label that I am aware of is the distinction between *specific ability* and *general ability* – these are used to refer to what I am calling freedom and ability, respectively.[20] I will not argue that my way of labeling the distinction is in anyway preferable, but it will at least be *uniform*. What is important here is to distinguish clearly

---

discussed here is abilities to engage with some particular thing or some particular kind, such as the ability to shake Obama's hand, or the ability to sculpt gold. These abilities might be thought to depend not only on the intrinsic properties of their bearers but also at least on the *existence* of Obama and of gold (and not merely of some person who is similar to Obama, or of some substance that is similar to gold). So condition (ii) may need to be revised in light of such cases. I conjecture however that any such revision would preserve *some* sort of intrinsicality constraint on the ability relation, while there is, as I suggest in what follows, *no* such constraint on the freedom relation.

[20] In, for example, Mele 2002. This terminology suggests what I think is plausible – that the ability relation is just a certain kind of generalization from the freedom relation. (van Inwagen makes a similar conjecture in van Inwagen 1983). I will not pursue this point here, though if it were true it would clearly pose a deeper obstacle to the reductions of freedom to ability that I canvass in the next chapter.

the freedom relation from the ability relation, and to recognize that the former relation will be the topic of this essay.

1.6 "Can" and ascriptions of freedom[21]

The modal auxiliary "can" does not always ascribe freedom to agents. But sometimes I think it *is* used to ascribe freedom to agents. Indeed, such ascriptions have already figured and will continue to figure in this essay. As I have said, I will not offer anything like a systematic account of the semantics of "can" and cognate phrases. Nonetheless, let me make some brief remarks about how I think of the relationship between the present essay and that semantic project.

The most notable consequence is the following. I have argued here that "can" is sometimes used to ascribe freedom, and I will argue in the next chapter that the freedom relation is irreducible to any other relation. If this is correct, then any adequate truth-conditional semantics for "can" will include the freedom relation among its constituents. Any account of what make such ascriptions true, that is, will have to make mention of freedom, for freedom is what these sentences sometimes ascribe.

This will make the semantics of "can" somewhat less simple than one might have hoped, in at least two respects. First, as I have already noted, it introduces complexity into the relations ascribed by "can": for "can" is used to ascribe relations other than the freedom relation. Second, it introduces complexity into the objects to which these relations are ascribed. For sentences including "can" may be true of both agents and non-agents (tools, for example), yet the freedom relation is one in which only agents stand.[22] It is thus I think a constraint on attempts to give a systematic semantics for "can" that they accommodate at least this degree of complexity.

This is all I will have to say about "can" and cognate phrases. There is sometimes a tendency in discussions of freedom to conflate the formal with the material mode. My concerns thus forth will be solely with the latter. I have claimed that "can" is sometimes used to ascribe the freedom relation to agents. The rest of this essay will be addressed to the nature of that relation.

1.7 A framework for freedom

One of the advantages of specifying the freedom relation at length is that it allows us to clearly and concisely pose a number of traditional problems about free will in terms of

---

[21] I will focus in this section on "can," but as I note in the text similar remarks apply to cognate phrases.

[22] Neither this point nor the previous one entail what I think is implausible, namely that "can" is genuinely *ambiguous*, so that the occurrences of "can" in "I can now raise my arm" and "Telescopes can detect distant planets" have nothing in common semantically. More plausible I think is that each of these expresses some common relation, such as restricted possibility, but that the correct semantics nonetheless requires making reference to the freedom relation. For example, in the case of restricted possibility, this would mean that the domain of the restriction in some cases could not be captured without appealing to the freedom relation.

the freedom relation. Let me say what those questions are and how I plan to address them in the remainder of this essay.

There are at least three questions to be asked about the *metaphysics* of the freedom relation:

(i) Is the freedom relation ever instantiated?
(ii) Is the freedom relation *reducible* to some other relation?
(iii) Does the instantiation of the freedom relation *entail* the falsehood of determinism?

Roughly corresponding to these are three questions about the *epistemology* of the freedom relation:

(iv) Do we have any *evidence* that the freedom relation is instantiated?
(v) Is the instantiation of the freedom relation *inferable* from some other facts?
(vi) Does our evidence for the instantiation of the freedom relation *justify* the belief that determinism is false?

The remainder of the essay can be outlined in terms these six questions.

In Chapter Two, I will argue for a *negative* answer to question (ii). The freedom relation is not reducible to any other relation. I therefore reject, *inter alia*, the traditional compatibilist proposal that facts about freedom reduce to facts about the counterfactual dependence of one's actions on one's choices.

In Chapter Three, I will argue for a *negative* answer to question (v). The instantiation of the freedom relation cannot be inferred from any other facts. In much the same way that Hume claimed it is never legitimate to infer an "ought" from an "is," I argue that it is never legitimate to infer a "can" from a "does."

In Chapter Four, I will argue for a *positive* answer to question (iv). Our faculties of bodily awareness give us evidence for the instantiation of the freedom relation, for our freedom to engage in various bodily actions is part of the *content* of bodily awareness. This experience normally constitutes, I argue, *perceptual evidence* for the instantiation of the freedom relation.

In Chapter Five, I will argue for a *negative* answer to question (vi). Our evidence for the instantiation of the freedom relation is not the right *kind* of evidence to justify us in believing the falsehood of determinism. I argue that the most defensible strand of traditional compatibilism is a negative answer to this question, and in that sense the present account vindicates what is true in compatibilism.

This leaves the traditional questions about the metaphysics of freedom – (i) and (iii) – unanswered. In the Conclusion, I argue that we have no grounds to answer these

questions beyond what was given in my answers to questions (iv) and (vi). But since these questions are distinct, it would require an unreasonable dogmatism to assume that the answers to them coincide. It may therefore be that, given the limits of our experience, these are questions that, for us, must remain open.

1.8 A note on "freedom of choice"

In sections 1.3 and 1.4, I took the relata of the freedom relation to be actions, and took actions to at least *include* basic movements of the body, such as arm-raisings. As I understand it, our freedom, if in fact we are free, is at least in part a freedom to *do* things, to immediately bring about changes in the world: lifting one's arm, for example. But this point is not universally accepted in writings on freedom. Sometimes it is said that the central question in this area is a question about freedom of *choice*: are we free to make *choices* other than the ones we in fact make? To frame the question in the way I have, according to this way of thinking, involves a confusion between freedom, which is a property of our *minds* or *wills*, and the various relations that hold between our thinking or willing in a certain way and the changes that obtain in the world in response to our so thinking or willing. Since this objection holds that there is an error in the very way of conceiving of the problem, I want to close this chapter by briefly responding to it.

It is true that when we speak of what someone is free to do, we often speak the language of volition. We say, of someone selecting from a variety of sandwiches on a menu, that it is *his choice* which one he has. This ordinary way of speaking is reflected in our philosophical terminology. When we speak of whether someone is free to do otherwise, we often ask simply whether he has "free will."

When we speak this way, I think we are speaking metonymically: we are speaking of a *part* of freedom, albeit a crucial part, in place of the whole. There are a variety of ways in which it could fail to be someone's "choice" what sandwich he has. *One* way is through a disorder in volitions, for example a compulsion against choosing to eat fish. Another way is through more mundane obstacles, for example the kitchen being all out of tuna. In this way, the freedom relation is much like the knowledge relation: it is a relation that is instantiated in virtue of both psychological facts about the agent to whom freedom or knowledge is ascribed *and* non-psychological facts about the world and the agent's position in it.

This is, I think, the intuitive way of conceiving of freedom. What grounds might someone having for claiming that it rests on a confusion? I can think of two ways, one more radical than the other. The more radical way holds that the present view is confused about the very idea of action. Actions do *not* include movements of the body. All actions are inner events, or what are sometimes called *volitions*. Since freedom is, all are agreed, a relation to actions, it is a confusion to think of events like arm-raisings as falling within its domain. For these events are not actions at all, but are rather consequences of the things we do.[23]

[23] This is the view defended in Hornsby 1980.

This way is more radical because it calls into question a piece of orthodox action theory that I have simply assumed, according to which bodily movements are at least sometimes among the things we do.[24] If this alternative view of action is indeed correct, then the present essay is misguided (how deeply it is misguided, and whether or not it could be "reconstructed" in terms of this purified action theory, is another question). I think this view of action is not correct, but I will not argue that here. I think we would need a more complete theory of agency than I will provide here to answer this sort of challenge. One must begin somewhere, and I am beginning with the natural – though not, as we are now pointing out, indubitable – thought that actions sometimes include bodily movements.

A less radical way does not call into question the view that actions at least include bodily movements. Its concern is rather local to the issue of freedom. This is the point already gestured at above. According to this way of speaking, the phrase "free will" should not be taken metonymically at all. For freedom is in the first place a property of the will. To speak of the freedom to walk is to confuse two quite separate issues: freedom of will, on the one hand, and the physical power to walk on the other.[25]

This way of objecting to the present proposal is, as I have said, less radical, but for the same reason it does not seem to be well-motivated. Once we have granted that *action* admits of both "inner" or psychological components *and* "outer" or non-psychological ones, it is not clear why we should deny that freedom is similarly composed. There do not at least seem to be any grounds here for accusing the intuitive view according to which we are sometimes, for example, free to raise our arms of confusion. It *may* be that all such talk is misleading and needs to be reconstructed within a framework where freedom is a relation in the first place to choices, but unless we adopt the radical view that all actions *are* choices, no such reconstruction seems obligatory. So I will proceed in taking at face-value the view that the things we are free to do include movements of our bodies, such as arm-raisings.

This way of viewing matters is not unimportant, for it will be central to understanding some of the arguments that will follow. In Chapter Four, I will argue that our faculties of bodily awareness – which I will be take to be ways of *perceiving* our extended selves – are what reveal to us what evidence we have for thinking that the freedom relation is instantiated. To be aware of one's arm, for example, involves, *inter alia*, an awareness of one's freedom to move that arm. If freedom were simply a property of our minds or wills, it would not be the sort of thing that could be so perceived. So I will come to argue that there is a deeper reason for rejecting the view that freedom is simply a property of our minds or wills. For we are *acquainted* with the freedom relation in the first place as a relation between ourselves and the spatially extended bodily movements that we are free to perform.

This point also bears on the limits of what is shown by the arguments that follow. The evidence delivered by bodily awareness is I think compatible with the truth of a variety of

---

[24] The *locus classicus* for this view is Davidson 1971: "We never do more [or, it should be added in the present context, less] than move our bodies; the rest is up to nature."
[25] This is the view of Albritton 1985.

theories of human psychology and of "the springs of action." In particular, it is compatible with theories according to which we are all like compulsives in the relevant sense, in which each of our actions is driven by desires so powerful as to be insensitive to our volitional control. If such a theory is true, then I do not think the present essay will have nonetheless shown that we at least enjoy "freedom of bodily movement," though we do not in fact enjoy "freedom of choice." For, if such a theory is true, then the freedom relation is simply *not* instantiated. And if that is true, then the evidence I will adduce in favor of the claim that we *are* free is all of it misleading. So the confirmation of such a theory would show that this essay constitutes an extended argument for a false conclusion. This is one way in which the claims I make here are vulnerable to the discoveries that psychology may yet make about us.[26]

[26] This marks a difference between the present argument and the "paradigm case" arguments usefully discussed in van Inwagen 1983.

The Possibility of Freedom

Chapter 2: Against Reduction

2.1 What is a reduction of freedom?

The project of traditional compatibilism has been closely bound to a subsidiary project: the project of giving a reduction of freedom. Informally, a reduction of freedom purports to show that facts about freedom *just are* facts of some other sort, for example facts about what someone would do if he chose to do so. In this section I will try to say somewhat more formally what a reduction of freedom would be.

This is not always done. As a result, discussions of what I will take to be proposed reductions of freedom are sometimes not discussed with the perspicuity that they demand. For example, discussions of the "conditional analysis" are often framed in linguistic terms, so that the question turns on whether or not "can" is in some sense "hypothetical" or "categorical."[27] But this question is only indirectly related to the question that we seem to be asking when we ask whether freedom can be reduced, for this question is in the first place a question about freedom itself, rather that the language in which we speak about it.

We therefore need a framework in which the metaphysical aims of the reductive project can be captured. I think the apparatus I have already introduced allows us to do this. In the previous chapter I introduced the freedom relation, a relation between agents and acts they do not perform whose canonical form is:

(FR) S is free to A

The task for a reduction of freedom can be defined in terms of (FR). It seems to be at least a *necessary* condition for a reduction of freedom that it find some other relation R such that:

(RED) S stands in R to A iff S is free to A

Any relation that did not satisfy (RED) would not plausibly be a reduction of freedom, for there would be actions such that some agent stood in that relation to those actions but was not free to perform them (or, conversely, was free to perform them but did not stand in that relation to them).

This necessary condition, however, is too easily satisfied. For there are clearly a number of relations that satisfy (RED). One is the freedom relation itself. But there are other relations as well, for example the relation of being free to perform an action *and being an agent*: this is a relation that an agent stands in to an action just in case he is free to perform it. So the freedom relation and what we might call *trivial transformations* of the freedom relation satisfy (RED).

---

[27] In, for example, Austin 1956.

This problem is one we can resolve as follows. It is necessary for a reduction of freedom that there be some relation R such that:

(i) R satisfies (RED), and
(ii) R does not include the freedom relation among its constituents

The second clause excludes the freedom relation and its trivial transformations from the domain of permissible relations. It captures the intuitive idea that a successful reduction of freedom should not be "circular": that it should *eliminate* any reference of freedom from those facts to which freedom is to be reduced.

It may be that even finding a relation which did satisfy (i) and (ii) would not be *sufficient* for providing a reduction of freedom. This would be the case if there is some further demand on a reduction beyond necessary coextension. Someone might hold, for example, that the properties of *being good* and *maximizing happiness* are necessarily coextensive, but that goodness nonetheless does not reduce to the maximization of happiness. But this point is moot, for I will argue that there is no relation that satisfies (i) and (ii). Since satisfying (i) and (ii) is plausibly a *necessary* condition for a successful reduction, it will follow that there is no reduction of freedom. We can therefore set to one side deeper concerns about the demands on reduction.

That is what a reduction of freedom would be. Let us now ask what motivates the search for a reduction of freedom in the first place. What *work* is a reduction of freedom supposed to do?

2.2 Three roles for a reduction of freedom

The project of offering a reduction of freedom has been pursued to a degree that may seem disproportionate to the mere intrinsic interest of such an analysis. Like similar reductive projects – for example, of normative and phenomenal properties – the main motivation for this project is largely extrinsic, deriving from the contributions that such a reduction might be thought to play in the pursuit of some larger end. In this section, I want to distinguish *three* roles that a reduction of freedom – that is, a relation that satisfies (RED) and does not include the freedom relation itself as a constituent – might play.

One role that a reduction might play is a role in adjudicating certain *arguments*. In the case of freedom, the project of reduction is typically though to have some bearing on the debate between the compatibilist and the incompatibilist. In particular, a reduction of freedom is typically thought to play some role in undermining the incompatibilist's arguments for his position. Let us call this the *dialectical* role of a reduction of freedom.

In its dialectical role, a reduction of freedom is used as follows. The incompatibilist offers arguments to show that freedom is incompatible with determinism. The compatibilist in turn proposes his reduction, which, if correct, is substitutable for every

occurrence of "freedom" in those arguments *salva veritate*. But the incompatibilist's arguments are unsound when the proposed reduction is substituted into them. Therefore the arguments are unsound in their original form, since the mere substitution of equivalent terms does not render an unsound argument sound, though it may make that argument *seem* to be sound.

There is an obvious problem with this strategy, however, namely that one man's *modus ponens* is another man's *modus tollens*. The incompatibilist may respond that since his arguments clearly *are* sound, and the arguments with the compatibilist's reduction inserted into them are *not* sound, this itself is a decisive counterexample to the proposed reduction. A good reduction must preserve all of the features of that which is to be reduced, including the features that it exhibits in the incompatibilist's arguments. It does not seem, then, that a reduction would do anything to advance the standoff between the compatibilist and the incompatibilist. In this respect, the dialectical role is not one that a reduction of freedom can be reasonably expected to fulfill.[28]

There is another role for a reduction of freedom, however, that is *internal* to the compatibilist program. The compatibilist's claim, namely that one may sometimes be free to do what one was predetermined not to do, is at least *prima facie* implausible. What motivates the compatibilist's claim is a larger methodological aim: to show that our ordinary thought and talk about freedom is not vulnerable to the discoveries that may be made by final physics. A reduction of freedom, if it is itself plausible, provides a way of realizing this aim. It might yield a theory of freedom that *explains* how freedom can be compatible with determinism. Let us call this the *ontological* role of a reduction of freedom.

Of course, anyone who already accepts the argument that freedom is incompatible with determinism may reasonably reject the reduction. The role of a reduction of freedom, in this respect, is not to *convince* the incompatibilist of the error of his ways. It is rather to render intelligible the compatibility of freedom and determinism to the compatibilist's own satisfaction. This is the way in which the dialectical and the ontological role come apart.

The potential of a reduction to play such a role is, I think, much of what explains its importance to the program of traditional compatibilism. In the remainder of this chapter, however, I will argue that the ontological role will have to go unfilled, since there *is* no reduction of freedom to be had. This raises an important question: what are the prospects for compatibilism once we have rejected the project of giving a reduction of freedom? What, in short, are the prospects for a *non-reductive compatibilism*? These questions have largely gone unasked in the free will literature because the project of offering a reduction of freedom and the project of defending its compatibility with determinism have been so closely bound to each other. But this tendency obscures the availability of the following claim: that freedom is irreducible *and* it is compatible with the truth of determinism. We will be in a position to assess the this claim only at the very end of this

[28] van Inwagen makes the same argument in van Inwagen 1983.

essay.[29] So, while the arguments of this chapter are I think fatal to the prospects of a reduction of freedom, they do not yet decide the prospects for compatibilism more generally.

Finally, there is a third role a reduction of freedom might play. In the next chapter, I will emphasize the difficulty of giving an adequate answer to epistemological questions about the freedom relation. Are we in a position to have justified beliefs about what we are free to do, and if so how are those beliefs justified? If a reduction of freedom were available, then it would yield straightforward answers to those questions. Let us call this the *epistemological* role of a reduction of freedom.

Here is one way in which the epistemological role might be filled. If freedom admitted of a conditional analysis, then the epistemology of freedom would reduce to a special case of the epistemology of conditionals, for to be justified in believing a claim about the freedom relation would just be to be justified in believing a conditional with a certain content. This would not make the epistemology of freedom *straightforward*, for the epistemology of conditionals is not itself a straightforward matter, but it might at least make it *tractable*.

Once we reject a reduction of freedom, these epistemological questions remain open, and in a way more difficult. For the epistemology of the freedom relation must be in a certain way *sui generis*: we cannot give an account of it *just* by giving an account of our justification for believing some claims about some other relation, for claims about the freedom relation are distinct from any such claims.[30] This is the problem that will animate Chapter Three, and to which my own positive proposal in Chapter Four will be a response. If this line of argument is sound, then the absence of a reduction of freedom has significant consequences for the form that an adequate theory of the freedom relation will take.

2.3 The conditional analysis[31]

A recurrent strain in traditional compatibilist thought has been that claims about freedom are in some sense equivalent to claims about what agents *would* do if their mental states were different. Though it is generally agreed that the conditional analysis is mistaken, I tend to think the ways in which it is mistaken have not been made adequately clear. Here I will try to do better. In this section I will state the conditional analysis as clearly as possible; the framework established in the previous chapter puts us in a good position to do this. In the next several sections I will consider and reject a couple of proposed revisions to the conditional analysis, and also consider the degree to which objections to

[29] In Chapter 5 and the Conclusion.

[30] Of course, for all I have said thus far it could be that though the freedom relation does not reduce to any other relation (as I will argue in this chapter), facts about the freedom relation might nonetheless be *inferable* from facts about other relations. But a central argument of Chapter 3 will be that this is not in fact possible: that facts about freedom are not only not *reducible to* but also not *inferable from* any other facts.

[31] I use the terms "reduction" and "analysis" interchangeably throughout, favoring the latter term in the next few sections since the phrase "conditional analysis" is so entrenched.

the conditional analysis of freedom turn on obstacles to conditional analyses more generally.

The conditional analysis, in our terms, claims there is a relation that satisfies (RED) and does not include the freedom relation as a constituent. This relation is a *counterfactual* one, a relation between mental states on the one hand and the actions he would perform if he had those mental states on the other. In one form, the conditional analysis claims that:

(CA) S is free to A just in case if S were to choose to A then S would A[32]

Proponents of the conditional analysis are not unanimous on *which* mental state occurs in the antecedent: some substitute, for example, "were to try" for "were to choose." I will presently consider and reject a revision of (CA) that turns on this sort of point. For now let us focus on this form of the conditional analysis, and later consider whether revisions to the antecedent can resolve any of the problems I will raise for (CA).

Counterexamples to (CA) are familiar.[33] Consider some agent with a bowl of tomato soup in front of him, equipped with a spoon and so forth. It is true of this agent that, if he were to choose to eat some tomato soup, he would eat some tomato soup. But it is compatible with this that the agent in question has a compulsion that prevents him from choosing to eat anything that is a particular shade of red, which is precisely the shade that this soup happens to be. This agent is *not* free to eat the tomato soup. Eating the tomato soup is not something that is in his power, or is among his options – in *just the same way* as eating this tomato soup is not among the options of someone who is gagged and bound. Yet this agent does stand in the counterfactual choice relation to eating his soup. So this agent is a counterexample to (CA). It is true that this agent would eat this tomato soup if he chose to, but he is not free to eat the tomato soup.

Does the problem with (CA) lie, as suggested above, in the mental state that we have chosen as an antecedent? Donald Davidson claims that it is when he argues that:

> The antecedent of a causal conditional that attempts to analyze 'can' or 'could' or 'free to' must not contain, as its dominant verb, a verb of action, or any verb which makes sense of the question, Can someone do *it*?[34]

Davidson therefore proposes to his own alternative to (CA), according to which:

> S can do A intentionally (under the description d) means that if A has desires and beliefs that rationalize x (under d), then S does x[35]

---

[32] Recall that a proponent of the conditional may be claiming something stronger than mere coextensionality, if mere coextensionality does not suffice for reduction. But the coextensional claim made by (CA) is at least *entailed* by the conditional analysis, so counterexamples to (CA) are also counterexamples to any stronger claim made by the proponent of the conditional analysis.

[33] The counterexample given in this paragraph is a variant on one given in Lehrer 1968.

[34] Davidson 1973, p. 68

But this proposal would seem to be subject to the counterexample just given. Let 'd' be 'eating a bowl of red soup'. If this agent had a desire to eat a bowl of red soup sufficiently strong to rationalize doing so, then he would eat it. But the problem with this agent is precisely that his actual psychology is incompatible with the existence of any such desire. So it is not true that this agent is free to intentionally eat this bowl of soup. So Davidson's proposed alternative to (CA) does not succeed. More generally, it appears that any revision along these lines will not succeed, since we can reconstruct this counterexample for whichever mental state may occur in the antecedent of the proposed conditional. This objection to (CA) does not turn on the particular mental state which its proponent happens to have selected.

What it *does* turn on is something that will not become clear until we have considered the problems with analyses that attempt to improve on (CA). But we may informally capture the problem with (CA) as follows. (CA) abstracts away from an agent's *actual mental states*. But an agent's actual mental states are, sometimes, precisely what renders him unfree. Thus (CA) overgenerates, counting agents as free to perform certain actions by removing precisely the conditions in virtue of which such agents are not free to perform those actions.

2.4 A strengthened conditional analysis

A natural thought occurs in response to the kind of counterexample that created problem for (CA). (CA) seems to face trouble when applied to agents who are psychologically abnormal in some sense or other, in particular those whose volitional capacities are defective. Why not simply add a *second* clause to the conditional analysis which rules out such agents? An agent is free to A just in case he would A if he chose to *and* he satisfies some further condition. The difficulty is to spell out this further condition in a way that does not itself invoke the freedom relation that is to be analyzed.

Christopher Peacocke has recently proposed an analysis along precisely these lines. Peacocke suggests that a given agent

> is free to A just in case:
>
> (a) he could (closeness possibility) try to A
>
> (b) he would A if he tried to.[36]

The (a) clause of this analysis is meant to play the role of ruling out agents who are in some sense or other psychologically abnormal. It does so without circularly appealing to the freedom relation itself. Instead, Peacocke appeals to a concept of "closeness." Peacocke takes closeness to be an intuitive notion, giving as a canonical example its role in the following case:

---

[35] Ibid., p. 73. Davidson ultimately rejects this analysis as well, but for reasons that are not specific to freedom, and which turn instead on a problem for the causal theory of action more generally (the so-called problem of "wayward causation").

[36] Peacocke 1999, p. 313.

> Suppose you travel on a train through the Channel Tunnel, and there is a fire in the engine. Suppose also that the only reason that the fire does not spread poisonous smoke through the ventilation system is that some baggage, which could easily have been placed in a different configuration, happens to set up a draught which diverts the smoke from the ventilation system. It is true to say of this situation that there *could* easily have been a fatal accident.[37]

The sense of 'could' in Peacocke's clause (a) is just the sense that it has in the description of this example – however exactly that sense is to be spelled out. Since this sense of "could" appears to be independent of the freedom relation that we are trying to analyze, Peacocke's analysis, if correct, would constitute a successful reduction of freedom. (More precisely, it would at least satisfy the criterion of extensional adequacy that is a necessary condition for any successful reduction of freedom).

But Peacocke's analysis is not correct. Consider the following case. A traveler on the train just described has a bowl of tomato soup before him. As before, if he were to try to eat the soup, he would. So he satisfies clause (b). Also as before, this agent has a compulsion against eating anything that happens to be the particular shade of red of this soup. But imagine that the following is also the case. There is a fire in the engine, yet fumes from the fire are diverted away from the passengers, as in Peacocke's example. Yet these fumes are not toxic. Rather, they contain anti-psychotic drugs that relieve people of their compulsions (the train company finds this calms passengers down during emergencies). There is a close possibility where these fumes enter the car, and therefore a close possibility where the agent is freed of his compulsion, and so (since that compulsion was all, let us assume, that prevented him from trying to eat his tomato soup) a close possibility where this agent tries to eat his tomato soup. So this agent satisfies clause (a): he could (closeness possibility) try to eat his tomato soup. So according to Peacocke's analysis, this agent as he *actually* is – unrelieved by the fumes and in the grip of his compulsion – is free to eat the tomato soup. But this agent is not free to eat the tomato soup, in *just the same way* in which the compulsive described in the previous section was not free to eat his soup. So this proposed strengthening of the conditional analysis fails.

I have not considered all the attempts one might make to strengthen the conditional analysis in this manner.[38] But the problem with Peacocke's proposal I think underscores an obstacle to any such strengthening, one already mentioned at the end of the previous section. The problem with the conditional analysis of freedom is not that it fails to handle this or that case of psychological abnormality. The deeper problem was that it

---

[37] Ibid., p. 310; italics mine.

[38] There is another proposal which I am not considering, one which looks superficially similar to a strengthened conditional analysis but which is not in fact a reduction of freedom at all. According to this proposal, one is free to A just in case one is free to choose to A and one would A if one chose to. This proposal takes the freedom relation as primitive, and so does not count as a reduction of the freedom relation in the terms of the present chapter. If there are problems with this proposal, they lie rather in the general methodological concerns about the priority of action to choice outlined in Section 1.8. Since those concerns are, as I grant there, indecisive, this sort of *non-reductive* proposal about freedom remains available even if we accept the arguments against reduction that I give in this chapter.

abstracted away from an agent's *actual* mental states, thus missing cases where an agent's actual mental states are precisely what render him unfree. An account like Peacocke's, which tries to avoid this problem by adding some *further* modal condition on an agent's mental states, is bound to fall to counterexamples that, while they differ in details, trade on the same underlying idea: that part of what determines what someone is free to do is what his actual psychology is.

2.5 A normative analysis

The counterexamples in the previous two sections have turned on agents whose choices are in one way or another insensitive to their *reasons*. The compulsive may have an excellent reason to eat his tomato soup, but is rendered unfree to act on this reason by his compulsion. This observation suggests another way of revising the traditional conditional analysis. Why not say that someone is free to perform some action just in case he would perform that action if he had a sufficient reason to?

As it stands, this proposal cannot be correct. It is a familiar fact that agents regularly fail to act on their sufficient reasons. A smoker may continue to smoke, for example, even though the reasons in favor of his quitting are decisive. Perhaps *sometimes* the addict's behavior is to be chalked up to a lack of freedom, but it would be an excessive rationalism to demand that things were always thus. For it is sometimes the case that we have a sufficient reason to act in a certain way, are aware of that reason, are free to act in that way, and yet fail to act.[39] And if that is correct, then the analysis of freedom in terms of sufficient reasons cannot succeed.

We may fix this problem at least by going extensional. Consider the following analysis:

(CA*) S is free to A just in case there could be a reason to A such that S would A for that reason[40]

This analysis makes the correct prediction about the compulsive who caused problems for the analyses of the previous two sections. There could be *no* reason to eat the soup such that he would eat the soup for that reason. This is because his compulsion makes him such that he will not eat the soup for *any* reason whatsoever. Other cases present problems for (CA*) however.

Let us say that an agent is steadfastly *ignorant* of any reason to A. It may be, for example, that I have sufficient reason to buy stock in Google. Yet I am sufficiently ignorant of the market in technology stocks that any reason I might have to buy stock in Google is one that I will not be aware of. So there is no reason to buy stock in Google

---

[39] Compare Davidson 1970.

[40] Compare "weak reasons-responsiveness" in Fischer and Ravizza 1998. This analysis, like Peacocke's, has a modal expression occurring in the analysis (the "could" in "there could be a reason"), but, as with Peacocke's, it seems the sense of this modal can be made explicit without invoking the freedom relation itself.

such that I would buy stock in Google for that reason. Yet I am nonetheless free to buy stock in Google. So (CA*) fails as an analysis of freedom.

We might try to revise (CA*) as follows:

(CA**) S is free to A just in case there could be a reason to A such that S would A for that reason if he were aware of that reason

This analysis avoids the Google counterexample that caused trouble for (CA*), for there is a reason to buy stock in Google such that I would buy it if I were aware of that reason, such as that I would double my money for doing so. But (CA**) now faces the same sort of counterexample that felled the traditional conditional analysis. Imagine that someone is such that he *cannot* become aware of any reason to eat tomato soup. Such an agent is not free to eat tomato soup, *in just the same way* that the compulsive considered in the previous sections was not free to eat tomato soup. (Indeed, we can imagine that compulsions *in fact* work by constraining the cognitions of agents rather than by constraining their volitions). Yet (CA**) predicts that such an agent *is* free to eat tomato soup. So (CA**) is subject to the same objection as (CA): that by abstracting away from an agent's actual mental states it abstracts away from precisely the conditions that can make him unfree.

Another way of trying to salvage what is true in (CA*) is as follows. Note that the objections to (CA) turned on the *sufficiency* of the analysis: being such that one would choose to perform an action if one chose to is not sufficient for being free to perform that action. But the counterexample to (CA*) turned on its *necessity*: being such that there could be some reason to perform an action such that one would perform it for that reason is not necessary for being free to perform that action, for someone may be free to (for example) buy stock in Google without meeting that normative condition. May we say then that (CA*), unlike (CA), at least provides a *sufficient condition* for freedom?

We may not. There are counterexamples to the sufficiency of (CA*) as well. Consider the following. An agent is imprisoned in a jail, the door of which is set to open just in case there is a fire (this is a provision for the safety of the prisoners). There could be a reason to leave the jail such that the prisoner would leave the jail for that reason – namely, that there is a fire. Yet the prisoner, as he actually is, is not free to leave the jail. So while the Google example was a counterexample only to the necessity of (CA*) for freedom, this case is a counterexample to its sufficiency as well. So (CA*) fails in both directions.

The problem for (CA*) represents, I think, a generalization of the problem raised for (CA) and for Peacocke's proposed revision of it. The problem for those analyses was that the actual facts about an agent's mental states may be part of what determines what he is free to do. (CA*) attempted to overcome this problem by avoiding reference to an agent's psychology and instead appealing directly to what an agent's psychology normally responds to, namely his reasons for action. But this revision does not avoid the problem. For changes in an agent's normative situation supervene on change's in his

non-normative situation, and *any* change in an agent's non-normative situation – be it psychological or not – has the potential to change the facts that determine what he is free to do. In short, the lesson of these past few sections has been that there are *no* facts about an agent that may not be connected, either directly (like psychological facts) or indirectly (like normative facts), to the facts that determine what he is free to do. Any analysis that abstracts away from the actual facts about an agent may therefore, under some circumstances, abstract away from the constraints on his freedom. This constitutes a quite general obstacle to the prospects for a conditional analysis of freedom, an obstacle that is shifted but not removed by the proposed turn to normative facts in (CA*).

2.6 Dispositions and the "conditional fallacy"

I have been arguing that the problems facing the conditional analysis of freedom arise from distinctive features of the object of analysis, namely the freedom relation. In this section I will consider a diagnosis according to which the problems arise from the *method* of analysis. In recent decades it has become clear that there are general obstacles to giving conditional analyses. These obstacles have arisen most noticeably in attempts to give a conditional analysis of *dispositions*, and it is dispositions that will be my focus in this section, but the problems appear to be significantly more general than this, and have even earned the name the "conditional fallacy."[41]

Might the problems we have encountered thus far merely be instantiations of this more general "fallacy"? If they are, then proposals designed to overcome these problems in the case of dispositions and other features of the world might also offer a way of giving a successful analysis of freedom. But I will argue that this diagnosis is incorrect. The problems that arise for giving a conditional analysis of dispositions are quite different from the obstacles involved in giving a conditional analysis of freedom. Despite their superficial similarities, the failures of these two projects in fact arise from distinct sources. It is unlikely, then, that attempts to overcome the "conditional fallacy" in the case of dispositions will be of any use in achieving an analysis of freedom.

Let us briefly review the obstacles to giving a conditional analysis of dispositions. Intuitively, something is soluble just in case it would dissolve if placed in liquid. More generally:

(CAD) x is disposed to R in C just in case if x were in C then it would R

Where 'R' is to be filled in by the relevant response, for example dissolving, and 'C' is to be filled in by the relevant conditions or stimulus, for example being placed in water. (CAD) constitutes, in our terminology, a *reduction* of the disposition relation. The disposition relation is a relation that holds between stimuli and responses. According to (CAD), this relation is instantiated just in case *another* relation, the counterfactual dependence relation, holds between some stimulus and some response.

[41] See Bonevac et al. 2006. As is suggested there, this label may be a misleading one, for it is not clear that the obstacles to various conditional analyses reveal any more general problem that deserves the name "fallacy."

This proposal falls afoul of the problem of so-called "finkish" dispositions. Consider an ordinary sugar-cube, something that is soluble if anything is. Now imagine that this sugar-cube is watched over by a protective genie. If this genie sees that this sugar-cube is about to be placed in water, we will alter its crystalline structure to that of quartz, a material that is *not* disposed to dissolve when placed in liquid. According to the relevant instantiation of (CAD), this sugar-cube is not soluble: it would not dissolve if placed in water. But, intuitively, this sugar cube is soluble. It is true that, under certain conditions, it will become something that is *not* soluble – namely quartz. But this is entirely compatible with the claim that, as it actually is, this sugar-cube is soluble. This is a paradigm case of a "finked" disposition. And it appears to show that (CAD) fails as a reduction of the disposition relation.[42]

As a first step towards understanding why the problems faced by the conditional analysis of freedom seem to be rather different in origin, consider an agent who is like this sugar-cube in the relevant respect. He is very strong, but whenever he chooses to lift a heavy object, a malicious genie watching over him changes his muscle structure so that he is no longer strong enough to lift that object. His freedom to lift a heavy object is, as it were, finked. This case is a counterexample to (CA) just in case our intuition is that this agent *is* free to lift the object, just as the sugar-cube *is* soluble. But I think that our intuitions here run in precisely the opposite direction. The agent, as he actually is, is *not* free to lift the object in question – lifting a heavy object is, in virtue of the genie's machinations, not an *option* for him. Here then is a crucial *disanalogy* between the conditional analysis of dispositions and the conditional analysis of freedom: the freedom relation is not susceptible, as is the disposition relation, to finking.[43]

The disanalogy between these relations is even more striking when we consider the so-called phenomenon of "masking."[44] Consider again an ordinary sugar-cube. Imagine that this sugar-cube, rather than being subject to a genie's counterfactual interventions, is simply wrapped in a plastic coating through which liquid cannot pass. It is not true that this sugar-cube will dissolve when placed in liquid – for its plastic wrapping will keep the liquid from permeating it. Nonetheless, this sugar-cube is soluble: wrapping a sugar-cube in plastic does not make it insoluble, but rather prevents its solubility from manifesting itself. Yet according to (CAD) this sugar cube *is* soluble. So here is another reason to think that (CAD) is false.

Consider now the analogous case for freedom. Let us say that a strong agent is prevented from manifesting his strength, for example by being bound by steel ropes. Now consider

---

[42] Conversely, we may imagine a cube whose solubility is "reverse-finked": a cube of quartz that, when about to be placed in water, is transfigured by a genie into sugar. This cube is *not* soluble, but (CAD) predicts that it is soluble. So (CAD) seems to fail in both directions. These examples and this terminology for them are due to Martin 1996.

[43] This agent does perhaps have the ability to lift the weight, and plausibly this *ability* is finked. But here is a place where the freedom relation and the ability relation, as defined in section 1.5, come apart.

[44] Johnston 1992 introduces such cases. Fara 2001 makes the point that cases of masking are the decisive ones against the conditional analysis, fatal even to sophisticated conditional proposals like that of Lewis 1998.

this agent confronting a heavy weight that he is normally capable of lifting. Here I think it is clear that he is *not* free to lift the weight – in virtue of the steel ropes, lifting it is not an option of him. Obstacles are paradigmatic instances of features that prevent agents from being free. So it is not that the agent is free to lift the weight but his freedom is "masked" by the ropes.[45] The agent is simply, given the situation he is in, *not* free to lift the weight. The freedom relation, unlike the disposition relation, is not susceptible to masking.

The problems for the conditional analysis of freedom are different, then, from those faced by the conditional analysis of dispositions. The cases that constitute counterexamples to the latter analysis – namely, the cases of finking and masking surveyed in this section – are not counterexamples to the former analysis. The counterexamples to the conditional analysis of freedom are rather different in origin. They arise from the fact that, as I have said, abstraction away from the actual features of a situation sometimes entails abstraction away from precisely those features of a situation that render an agent unfree. I conclude that, whatever the hopes for giving an analysis (conditional or not) of the disposition relation, we have no reason to expect that such an analysis will also yield an analysis of the freedom relation.

2.7 The priority of freedom to ability

So much for the conditional analysis. There is another way of attempting a reduction of the freedom relation that does not turn on reducing it to a conditional relation. Let us now consider that attempt. In section 1.5, I distinguished the freedom relation from the ability relation. As I remarked there, these relations clearly stand in some kinship to each other. The question is what exactly the nature of that relationship is. One relationship that would be relevant to our present concerns would be if the freedom relation were *reducible to* the ability relation. If that were true, then this might constitute a successful reduction of the freedom relation.[46] But, as I will argue here, this view is not true.

We have already remarked that standing in the ability relation to some action does not *suffice* for standing in the freedom relation to it. So if freedom is reducible to ability then there must be some further condition, in addition to having an ability, that constitutes freedom. In the literature one sometimes sees the remark that freedom is just ability *plus opportunity*. This is typically offered as a platitude, but from the present point of view it

[45] Again, it seems true that his *ability* is masked, but this points again to a disanalogy between the freedom relation and the ability relation. As a more general point, attempts to overcome problems about freedom by assimilating the freedom relation to the ability relation, and the ability relation in turn to the disposition relation, seem bound to miss the sort of points insisted upon here. I suspect that the failure to make these distinctions is fatal to recent "dispositionalist" approaches to freedom, such as Vihvelin 2007 and Fara ms.

[46] *Might* because the ability relation itself would remain unanalyzed, and it might be that the proper analysis of this relation in turn demands appeal to the freedom relation (as I suggested in Chapter 1), so that we would embroiled in a certain kind of "circle of liberty."

constitutes a proposed reduction of freedom, and as such something that deserves our careful attention.[47]

The first question to be asked is: what is it to have an opportunity? We have, I think, a sufficiently intuitive grasp of this notion for it to play its required role in an analysis of freedom. A speaker has the opportunity to speak French just in case his vocal cords are working, there is no gag in his mouth, perhaps there are no irresistible threats against the speaking of French, and so forth. We need not have an analysis of the notion of opportunity, I think, in order to consider this proposed reduction. As long as having an opportunity to perform some action is not partly *constituted* by standing in the freedom relation to that action, then the present analysis, if correct, would count as a successful *reduction* of the freedom relation.

So our proposal is:

(AO) S is free to A just in case S has the ability to A *and* the opportunity to A

But consider the following case. Some agent has the ability to speak French – it is his native language, his vocal cords are working, and so forth. He also has the opportunity to speak French – he is not wearing a gag, there are no coercive threats in place against French speakers, and so forth. But imagine that the following is also the case. A malevolent genie is watching over him, ready to ensure that, at the moment he *tries* to speak French, his ability will be destroyed – the genie will in a moment erase all of the linguistic skills he acquired in his childhood in France. It is nonetheless true that this agent as he actually is does have the ability to speak French. And, in our intuitive sense of opportunity, it is true too that he has the opportunity to speak French. So, according to (AO), he is free to speak French. Yet, in virtue of the genie's potential interference, he is not in fact free to speak French. So (AO) is false.[48]

Shall we deny that this speaker has the opportunity to speak French? This strategy risks saving the reduction only by making it non-reductive. If any circumstances that deprives someone of freedom involves, *ipso facto*, an absence of opportunity, than our account of opportunity will require that we appeal to the freedom relation itself. Our intuitive account of opportunity in terms of actually absent obstacles was not subject to this objection, but it is subject to the counterexample just given for precisely this reason. In short, it seems we can make our concept of opportunity strong enough to avoid the sort of counterexample just given only by invoking the freedom relation itself.

It *may* be that there is some account of opportunity that avoids all such counterexamples without invoking the freedom relation itself. But even if there were, (AO) would still

---

[47] Susan Wolf offers this sort of proposal in Wolf 1980, p. 101, though she is careful to say that she is offering this as a "characterization" of freedom rather than a "reductive analysis." It seems though that *if* this were true then it would be a perfectly good reductive analysis.

[48] We can describe the general form of this counterexample in terms introduced in the previous section. (AO) requires that an agent have an ability that is not in some intuitive sense "masked"; the problem on which this counterexample turns is that this ability may yet be finked.

face another problem, one already raised in the previous chapter. This is that someone may be free to perform some action without having the ability to perform that action at all. Recall the tennis player who lacks the ability to hit a serve but who benefits from the interventions of a benevolent genie who will alter the course of the ball so that he hits a serve no matter what he does with the racquet. Such an agent is free to hit a serve, in these circumstances, but not because he has the ability to hit a serve. He has no such ability. He is rather made free to hit a serve in virtue of the extrinsic and fortuitous interventions of his genie.

The previous example objected to the *sufficiency* of (AO) for freedom. This example claims that (AO) does not even articulate *necessary* conditions for freedom. And, if it is correct, then it poses what seems to be an insurmountable obstacle to any reduction of the form of (AO). For (AO) attempts to reduce freedom to ability, but this example shows the *independence* of freedom from ability. One may be free to perform an action without having the ability to perform that action.

So we should reject the priority of ability to freedom, which was the idea on which (AO) turned. We may yet have grounds for accepting *the priority of freedom to ability*. This would be the case if we could, as suggested in the previous section, reduce the ability relation to the freedom relation. Here is one form that such a reduction might take:

(ABLE) S has the ability to A just in case S is normally, in virtue of his intrinsic properties, free to A

A proper defense of (ABLE) awaits a better articulation of what exactly "normally" demands.[49] Perhaps this cannot be done non-circularly, in which case (ABLE) would fail as a genuinely reductive account of ability. There may yet be other problems besides. Nonetheless, this seems to me a promising account of the nature of the ability relation. It is more promising, at least, than the converse attempt to reduce freedom to ability, which we have just given reason to reject.

2.8 Freedom as an irreducible relation

Here is a conjecture. There *is* no reduction of the freedom relation. That is, there is no relation that satisfies (RED) and does not include the freedom relation itself as a constituent. This is, as I say, merely a conjecture. I do not have a proof that there is no such relation.[50] For all I say here there could be a relation which does satisfy these

[49] A parallel problem occurs in the literature on dispositions, and one can think of various attempts to overcome the problems of finking and masking as attempts to make clear what exactly "normally" demands. Indeed, a claim exactly parallel to (ABLE) may I think be true of dispositions:

(DISP) x is disposed to R in C just in case x normally, in virtue of its intrinsic properties, Rs when C

On this view, just as the ability relation is a certain kind of generalization of the freedom relation, so the disposition relation is that same kind of generalization of the stimulus-response relation.

[50] Or even a general argument that might be deployed against any proposed reduction whatsoever, as Moore's "open question argument" purports to be.

criterion, but which I have not been able to discover. But I think it is a *reasonable inference* to make given the obstacles we have encountered in our search for an analysis.[51] We should at least take seriously the possibility that freedom is an irreducible relation, one that cannot be defined in terms that do not themselves make reference, implicitly or explicitly, to the freedom relation itself.

I am hardly the first to endorse this possibility. Richard Taylor, having surveyed the various properties that may be expressed by the modal auxiliary "can," writes

> I conclude, then, that "can," in the statement "I can move my finger," does not mean what it ever means when applied to physical things . . . What else is meant by "can," in this case . . . is suggested by what was just said; namely, that whether or not I do move my finger is "up to me" or, to use a more archaic expression, is something "within my power." And this is, certainly, a philosophically baffling expression which I feel sure no one can ever analyze; yet it is something that is well understood.[52]

This remark is, I think, merely an alternative way of expressing what I am calling the irreducibility of the freedom relation: the "can" in the relevant utterance of "I can move my finger" is unanalyzable just in virtue of expressing the freedom relation. And other philosophers have made similar remarks. Why then is the irreducibility of freedom not then taken more seriously as a substantive view on the metaphysics of freedom, even among philosophers otherwise sensitive to the limits of the project of reduction?

I suspect this is because philosophers, like Taylor, who have endorsed the view that freedom is irreducible have gone on to endorse a further view that many philosophers regard as intolerable, namely *incompatibilism*. One can see how this inference might be thought to follow given certain further substantive assumptions, for example that there is no way of resisting the incompatibilist's argument but by offering a successful reduction of freedom – that is, such arguments are resistible only if there is a reduction of the freedom relation fit to play what we have called its dialectical role. But we have already seen that this assumption is questionable. In any case, we can at least *distinguish* the question of whether freedom is reducible from the question of whether it is compatible with determinism. It *may* be that an argument can be constructed showing that the claim that freedom is irreducible leads inexorably to the conclusion that it is incompatible with determinism. If that is so then we can decide whether or not that conclusion is an acceptable one, and if not whether we need to revisit the case for freedom's irreducibility. But for now this possibility is simply being raised prematurely, for all I take myself to have done here is to establish a *prima facie* case for the irreducibility of freedom.

Where do we go from here? I think the best way to make progress at this point is by setting to one side these contentious metaphysical questions and focusing instead on an epistemological one. I said above that a reduction of freedom would have an *epistemic*

---

[51] Compare Williamson 2000, chapter 1, which makes what I take to be a similar argument about the knowledge relation.
[52] Taylor 1961.

role to play: it could explain how we are justified in believing claims about the freedom relation. If freedom is irreducible, then this role must go unfilled. This leaves open the following question: can we give an account of how we are *justified in believing* claims about freedom, once we acknowledge that freedom is irreducible? In the next chapter, I will draw out the difficulty of this question, which arises for the compatibilist and the incompatibilist alike. In the chapter after that, I will propose an answer to it. Only *then*, I think, will we be in a position to thoughtfully consider the fraught question of determinism.

The Possibility of Freedom

Chapter 3: Against Inference

3.1 Skepticism about freedom

With regard to any set of claims, one may ask what kinds of evidence support those claims. This question may be posed in a skeptical mode, challenging us to produce evidence for claims which seem to be widely but, says the skeptic, groundlessly accepted. This sort of skeptic has been largely absent from discussions of freedom, which is in a way surprising, for he has a compelling challenge to make. In this chapter I will present that challenge and reject some ways of responding to it.

Skepticism about freedom can usefully be thought of as a special case of a more general phenomenon, which we might call *skepticism about the modal*. The position is motivated by a very simple thought. We are acquainted, in the first place, with what is, rather than with what *could* or *must* be. So we are not in a position to know, or to have justified beliefs about, what is not actual. Nor does the testimony of others help us any, for they are our worldmates and so also acquainted, as are we, only with what is actual. Any claims about what could be or must be therefore cannot be anything but mere conjecture.

Skepticism about the modal comes in a variety of forms, more or less familiar and more or less compelling. As I have said, my focus here will be on skepticism about freedom, which is not particularly familiar but which is, I think, compelling. According to this view, we know in the first place only what we do and refrain from doing. The skeptic about freedom asks: what justification could we have for believing, of the actions we refrain from doing, that some of them are ones that we are free to do, and that some of them are not?

As there are a variety of subject matters about which one may be skeptical, so there are a variety of modes in which the skeptical challenge may be posed. One very ambitious kind of skeptic challenges us to provide arguments that convince him, beyond any reasonable doubt, that our position is the correct one. Another kind of skeptic demands only that we justify our position to our own satisfaction, employing canons of reasoning in which we are antecedently confident.[53] It is this *latter* skeptic with whom I will be concerned here. Though his challenge is modest, it proves, as I have suggested, a difficult one to meet.

In the previous chapter we canvassed one way of responding to the skeptic about freedom. A reduction of the freedom relation, if one were available, would yield the beginning of a proper account of its epistemology. For if claims about freedom were just claims about some other relation, then the epistemology of freedom might in turn reduce

[53] Compare Pryor 2000 on the "ambitious skeptical project" and the "modest skeptical project."

to the epistemology of that relation.[54] For example, if some form of conditional analysis were true, then an account of our justification for believing claims about freedom could be told in terms of our justification for believing certain conditionals. Of course, this account would be subject to its own skeptical challenge, but this would not be a *special* challenge about the epistemology of freedom. It would rather be a more general challenge about the relation to which the freedom relation reduces.

But there is no reduction of the freedom relation, and so the epistemology of freedom must be *sui generis*. There is nothing fit to play what we called the epistemic role of such a reduction. This is one reason why the skeptical challenge about freedom is especially compelling.

A second reason will emerge over the course of this chapter. The best response to certain kinds of skepticism about the modal, such as skepticism about lawhood, involves a certain kind of broadly inferential reasoning. Our best explanation of the world and its workings, says one sort of anti-skeptic about lawhood, postulates that there are laws. In this chapter I will show that such a response is not available in the case of freedom. Freedom does not figure in such explanations, and so is not the sort of thing for which we have broadly inferential justification. This is the epistemic cash-value of the thought that "the freedom discovered in reflection is not a theoretical property which can also be seen by scientists considering the agent's deliberations third-personally and from outside."[55] And this the second reason why skepticism about freedom is an especially compelling strain of skepticism about the modal.

3.2 An inductive argument

Perhaps the best place to begin to answer the skeptic about freedom is by considering *action*. We have, after all, ample evidence that we ourselves *act*. And this evidence is not solely introspective: we can also get evidence that others have acted. Perhaps *this* evidence is sufficient to ground our belief that we and others are free to act in ways other than the ways that they actually act.

The best development of this idea that I know of occurs in an article by Keith Lehrer.[56] Lehrer suggests that we could run a quite simple *experiment* that yields evidence that a given agent was free to perform an action. The crux of the experiment is this:

---

[54] Though it might not, for the reduction of freedom might not be *transparent* to us. So it might be that the epistemology of the freedom relation, at least for creatures like ourselves, involves getting justification directly about the freedom relation itself, rather than about that to which it reduces.

This point is especially salient if there is an *infinitary* analysis of the freedom relation, one which are not capable of making explicit or perhaps even understanding. (This possibility was pointed out to me by Eli Hirsch). Since my anti-reductive arguments in the previous section appealed only to the apparent unavailability of any explicit reductive analysis, they were not sufficient to rule out the possibility of such an analysis. But even if there is (unbeknownst to us) such a reduction of freedom, the epistemic point of this chapter still goes through. Given the limits of our comprehension, the epistemology of freedom must be, for us, *sui generis*.

[55] Korsgaard 1996, p. 96. As will become clear, I would say *perception* where Korsgaard says "reflection" and *bodily awareness* where Korsgaard says "deliberations."

[56] Lehrer 1966.

> Suppose that we instruct our subject to heed or not heed our instructions as he wishes, and ensure that the condition of the subject, as well as the circumstances in which he is placed, are those we have found to be most propitious for arm-lifting. Moreover, suppose that we watch him lift his arm, then avert our eyes for a moment, and, subsequently, see him lift his arm again . . . [then] we would have sufficient empirical evidence to support the hypotheses that the agent could have lifted his arm during that brief period when we did not see him lift his arm . . . In fact, even if we do not avert our eyes but see that he does not move his arm at the time in question, this in no way detracts from our evidence. Under the conditions we have imagined, the fact that our subject does not lift his arm need provide no evidence whatever to support the hypothesis that he cannot lift it.[57]

Let us consider the variant case with which Lehrer closes this passage, in which we do observe that the agent does not raise his arm at the time in question.[58] This experiment, if successful, would indeed silence *one* kind of skeptic about freedom. It might not satisfy the skeptic who demanded *certainty* that we were free to do otherwise. But it would at least satisfy the more modest skeptic who demands only that we have some evidence, adequate by the standards of ordinary empirical reasoning, that the freedom relation is sometimes instantiated.

The success of Lehrer's experiment, I think, turns on the soundness of an argument from observations of action to a certain claim about freedom. We can make this argument explicit as follows:

(1) We have justification for believing that the agent raised his arm on certain occasions [namely the two occasions when we saw him raise it]
(2) We have justification for believing that, on those occasions, he was free to raise his arm
(3) Since we have justification for believing that he was free to raise his arm on those occasions, we have justification for believing that he was free to raise it on other, relevantly similar occasions, including ones where he does not raise his arm[59]
(4) The moment when we saw that he did *not* raise his arm is such an occasion
(5) We have justification for believing that the agent was free to raise his arm on that occasion even though he did not raise his arm on that occasion

---

[57] Ibid., pp. 181-182.

[58] I focus on this case for the sake of simplicity. But the criticisms I will make go through equally well I think against the case where we *avert* our eyes during the moment in question, and so do not see whether or not the agent lifts his arms. For the evidence allegedly acquired in this case is somewhat weaker: instead of learning that an agent is free to raise his arm when he *does not* raise it, we would learn only that an agent is free to raise his arm when he *might not* have raised it. I argue here that this experiment fails even when our evidence is *strongest*, and so *a fortiori* in this variant case where our evidence is somewhat weaker.

[59] Lehrer's article includes a careful discussion of the conditions of relevant similarity, which he calls "temporal propinquity," "circumstantial variety," "agent similarity," and "simple frequency"; see pp. 177-181.

(C) We have justification for believing that this agent is sometimes free to perform actions he does not perform

If this argument were sound, then I think it would be true, as Lehrer claims, that we have inductive evidence for believing that this agent is free to perform actions that he does not perform. And since this experiment is merely an idealization of our normal epistemic position with regard to the actions of ourselves and others, we would have justification for believing that agents quite generally are often free to perform actions that they do not perform. But, as I will argue in the next section, we have no such evidence, because this argument is not sound.

3.3 Why this argument fails

This argument, I think, trades on an equivocation. There are two things that might be meant by the phrase 'free to' as it occurs in this argument. On one reading, premise (2) is true, but premise (3) is false. On the other, premise (3) is true, but premise (2) is false. But there is no reading of this argument on which both premises are true, so the argument is unsound, and Lehrer's "experiment" fails. Let me explain.

Let us take the second reading – the one on which (3) is true and (2) false – first. This is the reading which follows my usage of 'free to' thus far in this essay. On this reading, an agent is free to perform some action just in case he stands in the freedom relation to that action. I think it is unobjectionable that, on this reading, premise (3) is true. We should grant that for any property whatsoever, evidence that is instantiated on one occasion is evidence that is instantiated on another relevantly similar occasion, unless there is some independent reason to suspect that this is not the case. To deny *this* would to be a skeptic about induction, a position that is out of place here, where we are not doubting the accepted standards of scientific reasoning. So we should accept what premise (3) says on this reading, which is just that evidence for the instantiation of the freedom relation on one occasion is evidence for its also being instantiated on other, relevantly similar occasions.

On this reading, however, premise (2) is false. We do not have any justification for believing that the freedom relation is instantiated on the occasions we observed the agent raise his arm. And this is not because of some epistemic limitation on our part. Rather, it is because of a more basic metaphysical point. The freedom relation is a relation between agents and acts they do *not* perform.[60] It is therefore not even there to be observed on the occasions where the agent *does* raise his arm. So on this reading of 'free to', on which is just expresses the freedom relation, premise (2) is false.

Let us then take the first reading, on which premise (2) is true but (3) false. On this reading, which is perhaps the more natural one, 'free to' does not express the freedom relation. It rather expresses the *disjunction* of the freedom relation and the action relation. An agent is free to perform an action just in case he stands in the freedom relation to that action *or* the action relation to that action. On this reading, premise (2) is

[60] See Chapter 1.4 for a defense

true, for we have evidence that the agent does perform the action, and so that he is free to perform the action in this sense.

But on this reading premise (3) is false. The relation expressed by 'free to' is a paradigmatically disjunctive one, and so not one that is fit for "projection" into other contexts in the way that premise (3) demands. Our only evidence that the agent is free to raise his arm in this context is that he stood in the *action relation* to raising his arm in some other, relevantly similar, context. But we have evidence that the agent does *not* stand in the action relation to raising his arm in this context. We therefore have evidence that this agent is free to raise his arm in this more inclusive sense of 'free to' *only if* we have *independent* evidence that he stands in the freedom relation to raising his arm. The observations cited in premises (1) and (2) provide no such evidence. So premise (3), on this reading of 'free to', is false.[61]

The reasoning here may seem sophistical. It will seem especially so if one finds the argument given in the previous section compelling. I confess that things do not seem so to me. I am inclined to regard that argument itself as sophistical – for it would make it all to easy for an "experiment" like Lehrer's to yield evidence that agents are free to do perform actions they do not perform, which does not seem to be the sort of question that can be decided by *that sort* of evidence. The point made here offers a *diagnosis* of where that argument goes wrong. Since the point about the division of the freedom relation and the action relation strikes me as independently plausible, I am inclined to therefore reject this argument on the grounds just given.

This point is a crucial one, and someone who does not accept it will not go along with much of the argument that follows. It is therefore worth dwelling on. One way of elucidating what is at issue here is to speak in terms of *natural* relations.[62] I hold that the freedom relation and the action relation are the natural relations in the neighborhood. The semantic value of the English phrase 'free to' is a relation that ought to be dispensed with when we are viewing things from the ontological point of view, since it is a disjunction of the properly natural relations. My opponent, however, may respond that it is *this* relation, the one that I am calling disjunctive, that is the natural one. According to him, the relations that I am calling the natural ones are simply two modifications of this relation: the freedom instantiated when one acts and the freedom instantiated when one does not. How can we decide which of us is correct?

I have already adduced some considerations that bear on this question. For example, as I noted in Chapter 1.4, my way of viewing matters captures the idea that the freedom relation is that about which the compatibilist and the incompatibilist disagree. This consideration is, admittedly, not decisive. There is however another way of thinking

---

[61] The general schema of the reasoning appealed to in this paragraph is as follows. Let us say we have evidence for the obtaining of property P in some context C. We therefore have evidence for the obtaining of the disjunctive property P v Q in context C. We therefore have evidence for the obtaining of P v Q in some relevantly similar context C'. But let us say that we also have evidence that ~P in context C'. Then we have *no* evidence (at least not on these grounds) that P v Q obtains in context C'.

[62] In the terminology of Lewis 1983.

about the question of "naturalness." On this way of thinking, the question of which relations are the natural ones is to be decided much as it is in the natural sciences, partly by seeing how fruitful the results are when we *take* these relations as natural. If we eventually find ourselves in a blind alley, then it may be that we have to revisit the question of which relation is natural. But, as I will try to show in what follows, this is not so. Making explicit the distinction between the action relation and the freedom relation allows us to articulate a coherent and plausible view of agency. And that is itself evidence that it is these relations, rather than what I am calling their disjunction, that are the natural ones.

Before proceeding, let me note two criticisms of Lehrer's argument which I am *not* making. I have mentioned these in passing, but it is worth rejecting them explicitly and at some length since it is easy to confuse my criticisms with these distinct criticisms, ones which I regard as illegitimate.

The first way of criticizing this argument is by embracing *skepticism about induction*. On this way of criticizing this argument, premise (3) is false because we have no grounds for projecting freedom from cases where it is observed to cases where it is not. I reject this criticism. I am happy to grant, and find it plausible, that *if* we had grounds for believing that the freedom relation were instantiated on some occasion, then we would also have grounds for believing that it would be instantiated on some other, relevantly similar, occasion. I simply deny that Lehrer's "experiment" gives us *any* grounds for believing that the freedom relation is ever instantiated in the first place.

The second way of criticizing this argument is by embracing what we might call the *hiddenness of freedom*. On this way of criticizing the argument, premise (2) is false because freedom is not the sort of thing that can be observed. We can observe action and inaction, but freedom is not the sort of thing that shows up in our experience of the world. I reject this criticism too. My point in rejecting premise (2) above was a metaphysical, rather than an epistemic, one: the freedom relation was not *unobservable* but simply *not there* to be observed in the case where the agent raised his arm. Indeed, as will become clear, the denial of the hiddenness of freedom will be central to the project of this essay. I claim that freedom *is* the sort of thing that presents itself in experience, though not in the way that the argument requires. This argument requires that freedom is observed third-personally when we see other agents acting, while I will argue that freedom is observed first-personally in circumstances where we ourselves are *not* acting.

3.4 Inference to the best explanation

I argued in the previous section that a certain argument that we have evidence that agents are free to perform actions that they do not perform – that is, that the freedom relation is sometimes instantiated – fails. This argument was a broadly inferential one. More specifically, it relied on enumerative induction on agent's *actions* to generate evidence about agential *freedom*. I think, for the reasons just given, that such induction generates no such evidence.

It does not follow that *no* broadly inferential reasoning can offer evidence for the instantiation of the freedom relation. And this is not merely because of the abstract possibility that there is some line of reasoning that we have overlooked. There is rather a very specific kind of inferential reasoning that is widely accepted as a way of gaining justification about empirical matters of fact. This is *inference to the best explanation*.[63] This challenge is especially pointed because inference to the best explanation seems a plausible way of gaining justification about matters that are not establishable by enumerative induction alone and are subject to their own variety of skepticism about the modal – claims about *lawhood*, for example. Might not claims about freedom also be justified by the inference to the best explanation?

Over the next several sections I will try to make the case that they are not. This will be a relatively brief answer to a difficult question, and as such it is not decisive; I believe it at least constitutes, however, a *sketch* of how freedom might be dispensable from the explanatory point of view. Coupled with the claims I defend in the next chapter, it is part of a sustained defense of a view on which claims about freedom are "epiphenomenal" in the sense that they do not show up in the third-personal explanations of the sort offered by the sciences, while at the same time being claims that each of us has first-personal justification for believing.[64]

Let me explain two ways in which our question is a difficult one, and how I propose to respond to those difficulties. First, there are a variety of things that might be meant by the phrase "the inference to the best explanation," and it is no easy matter to say what the best reading of this phrase is.[65] My response to this difficulty will be as *ecumenical* as follows. The arguments will not turn on any particularly contentious view of, for example, inferential virtues such as simplicity. If successful, my argument will show that explanatory concerns give us *no* justification for believing that the freedom relation is instantiated under *any* reasonable construal of explanation and of what it is for an explanation to be a good one.

The second difficulty is the following. The freedom relation might in principle figure in the explanation of many different phenomena. How can be sure that, just because the freedom relation is not needed to explain some particular phenomenon, that it is not nonetheless needed to explain some *other* phenomenon? We cannot, of course, be sure of any such thing. But we can I think make a reasonable conjecture about the kinds of phenomena that the freedom relation is most likely to explain. These are *actions*. An appeal to the freedom relation, if it explains anything at all, can be expected to explain why agents act as they do. It would be at least surprising if the freedom relation turned out to be entirely dispensable from the point of view of action explanation yet necessary to explain some phenomenon other than action. This is because action and freedom are so intimately connected, even if it is difficult to say what exactly that connection is. So I will respond to the second difficulty, the diversity of potential explananda, by restricting our attention to action explanation.

---

[63] See Harman 1965.
[64] I develop this point at further length in the Conclusion.
[65] See Lipton 1991.

In short, I am making the following argument:

(1) If the freedom relation figures in any explanation at all, then it figures in action explanation
(2) The freedom relation does not figure in action explanation
(C) The freedom relation does not figure in any explanation at all

I have just offered a brief defense of premise (1), and will turn to a possible objection to it below.[66] In the next section, however, I will offer a defense of premise (2).

3.5 Action explanation

Let us begin with the account of action explanation given by Donald Davidson.[67] On Davidson's view, an action is an event with a certain kind of cause. More specifically, an action A is an event caused by an agent's desire for some state of affairs F and his belief that his Aing will bring it about that F.[68] There is far more to be said about Davidson's account and the criticisms of it, but I know of no convincing case against the basic form of this account.[69] So I hold that Davidson's account is essentially correct.[70]

All of this is, of course, contentious. Rather than offering an extended defense of Davidson, I want to focus our attention on a criticism of Davidson's view that lies at a somewhat deeper level. I think that this criticism is a valid one, not in the sense that it offers a *counterexample* to Davidson's theory but rather that it underscores an aspect of

---

[66] In Section 6. Another interesting objection to (1) runs as follows. It is often thought that there is some connection between freedom and *moral responsibility*, and that this connection is an *epistemic* one: that our confidence in our judgments about moral responsibility is what gives us grounds for believing that we are free to do otherwise. (This is the positive argument for freedom in van Inwagen 1983). It is generally thought (for example by van Inwagen) that the relation here is a *deductive* one, that moral responsibility *entails* the freedom to do otherwise. But one might also take this relation to be an *explanatory* one: that is, the best explanation of the facts about moral responsibility invokes the freedom relation. And if that were true, then this would be a compelling objection to (1).

One difficulty with developing this suggestion is that we lack an account of what it is to explain broadly normative phenomenon (which I take to include facts about value and rightness quite generally as well as facts about moral responsibility), and to what degree such explanations give us grounds for believing in the entities that figure in them. Absent such an account, I find it difficult to make precise the idea of explaining facts about moral responsibility, and so difficult to assess this objection to (1).

[67] Originally in Davidson 1963.

[68] Ibid., p. 5.

[69] The best discussion of these issues that I know of occurs in Smith 1998. It may be useful to divide counterexamples to Davidson's theory into two kinds. The first kind are the various purported counterexamples defended in Hursthouse 1991 and successfully responded to, I think, in Smith's article. The second kind are cases of "wayward causation," which Davidson himself (in Davidson 1973) seems to regard as insoluble. What this latter kind of case seems to call for is not that Davidson's theory be rejected, but that it be supplemented with a certain qualification: the desire and belief must not only *cause* a certain action, but must cause it *in the right way*. Whether this qualification can be reduced to other terms (as is claimed, for example, in Peacocke 1979) is, I think, an open question.

[70] So I think that the *action relation*, unlike the freedom relation, may admit of a successful reduction (modulo the issue of "wayward causation" raised in the previous footnote); in a particular, it may be reduced to a special case of the *causation relation*.

agency that has been *left out* of Davidson's theory. This criticism turns essentially on the issue of freedom. Thinking through it will help us arrive, I think, at a clearer understanding of the topic that is at issue in this section, namely the relationship between action explanation and the freedom relation.

What is the criticism? Consider an informal characterization of our self-conception as agents. We regard ourselves as having various *options* before us, and as performing one of those options in virtue of the considerations in favor of it, and the considerations against its alternatives. This informal characterization includes at least three kinds of entity: an *agent*, some *considerations* (or, in another terminology, *reasons*) for and against acting in certain ways, and a variety of *options* (or, in the terminology introduced here, actions to which that agent stands in the *freedom relation*).[71] We might call this informal characterization of agency the *deliberative perspective*.[72]

The criticism is that *none* of these entities show up in Davidson's account. If a desire-belief pair constitute a *sufficient cause* of action, then there is no need to appeal in the explanation of action to options, or reasons, or even to an agent. So Davidson's account would seem to render all of these entities, and the deliberative perspective quite generally, *epiphenomenal* from the point of view of action explanation. And, according to the objection, this is simply not plausible.

There are two broad ways of responding to this criticism. On the one hand are views which take this criticism to be decisive against Davidson's account, and sketch alternative positive views on which agents, reasons, and freedom do have an explanatory role to play. The most prominent theories here are the *agent-causal* theories of the sort defended by Roderick Chisholm and Timothy O'Connor.[73]

On the other hand are views which do not take this criticism to be decisive against Davidson's account, but which recognize that this criticism demands an account of where these missing entities "fit in" to the picture of agency sketched by Davidson, an account

---

[71] We must be careful here. From the point of view of deliberation it is not quite right to say that all and only actions to which an agent stands in the freedom relation are those about which an agent deliberates. For one thing, an agent may have false beliefs about what he is free to do. But even once we relativize matters to an agent's beliefs, there is a problem. If an agent believes he stands in the freedom relation to an action, then he believes he will *not* perform that action, since the freedom relation is a relation to actions one does not perform. And an agent does not deliberate about what he believes he will not do. I think the proper answer to this problem is to introduce a degree of epistemic possibility. An agent deliberates about an action only if he believes that he *might* stand in the freedom relation to it. That is, for all he knows, it might be the case that it is an option for him that he will not perform.; it is, in short, something that he may either do or leave undone. Such actions are the proper objects of deliberation.

[72] Following Smith 1994. This term has the misleading implication, which Smith himself rejects, that this perspective is a "standpoint" (in the sense of Korsgaard 1996 and Bok 1998) which is not capable of conflicting with the "standpoint" of the natural sciences. Like Smith, I deny this. The deliberative perspective implicitly involves claims about *what there is*, and these claims are the sort of thing which we can be true or false *simpliciter*. So this elucidation of the *content* of the deliberative perspective leaves open the question of whether this perspective is veridical or illusory.

[73] See Chisholm 1976 and O'Connor 2000.

that Davidson himself does not offer. This is a challenge that has been taken up in the case of agency by David Velleman, and in the case of reasons by Michael Smith.[74]

My response to this criticism is of the second kind. Like many, I find agent-causal accounts unsatisfactory, both because we have insufficient evidence for the existence of such causation and because it is not clear that, even if there were such causation, it would be capable of doing the work which its proponents demand of it.[75] So we should not reject a broadly Davidsonian account on these grounds. But we do, I think, need to supplant it. Unlike Velleman and Smith, who purport to reduce agency and normativity respectively to the materials of belief-desire psychology, I do not think that the way to supplant Davidson's account is by offering such a reduction. For, as I argued in the previous chapter, there is no reduction of the freedom relation, be it to the materials of belief-desire psychology or to a more robust ontology.

My approach here is rather different. First, I will make explicit how the freedom relation fails to do any explanatory work in Davidson's account. Insofar as we accept Davidson's account of action explanation, this will constitute a defense of premise (2) above. Then, in Chapter 4, I will explain how we can be entitled to believe in facts about the freedom relation even though they do not show up from the explanatory point of view. Taken together, these argument will I hope establish how we can accept Davidson's account of action explanation while at the same time making room for the freedom relation in our theory of agency.

Let us begin with the explanation of actions that we in fact perform. As we have noted, a desire-belief pair, conjoined in the right way, are sufficient to explain on Davidson's account why an agent acted as he did. There is no need to appeal to a further fact of freedom to explain his action. Indeed, if the account defended here is correct, then there is no further fact to be appealed to. For the freedom relation is a relation between an agent and acts he does not perform. The relation that holds between an agent and acts he does perform is the action relation. And that is precisely the relation of which Davidson's account provides a reductive explanation.

It may be objected that I am wrong about this, that the natural relation in the neighborhood is the one that I am calling the disjunction of the freedom relation and the action relation. If this were true, then it would cast doubts on the anti-inductive argument that I made above in Section 4. But it is not clear that it would help to find a role for

---

[74] See Velleman 1992 and Smith 1994.

[75] See the second half of van Inwagen 2000 for a clear exposition of the latter problem. As for the former problem, the claim of "insufficient evidence," my position is somewhat delicate. Unlike many critics of agent-causation, I am willing to accept a central contention of its proponents: that our experience of agency is capable of giving us warrant for claims about agency. The argument of Chapter 4 will turn on precisely such a claim. I deny, however, what agent-causalists also need for their epistemic claims to go through, namely that this experience provides sufficient evidence for revising the account of the world offered by the natural sciences. As will become clear in Chapter 5, I deny that the warrant acquired in our experience of agency *transmits* in the way that would be required to support such a revision. In short, we might say that I grant *more* epistemic weight to the experience of agency than do other critics of agent-causation, but *less* weight than the proponents of agent-causation themselves do.

freedom in action explanation. We would still not need to appeal to freedom to explain why an agent acted as he did; for again, the Davidsonian explanation is entirely adequate to explain this. Indeed, on this construal of freedom, the explanation goes the other way around. Rather than freedom figuring in the explanation of why an agent acted as he did, the explanation of why the agent acted as he did figures in the explanation of why he was free to perform it. For on this broad construal of freedom, acting in a certain way is sufficient for being free to act in that way, and so to explain why an agent acted as he did is also to explain why he was free to perform it.

Another way of trying to find room for the freedom relation (taken, as I have argued that it should be, as a relation holding only between agents and acts they do *not* perform) in action explanation proceeds from the thought that all explanation is *contrastive*. A full explanation of why someone acted that he did includes an explanation of why he acted in that way rather than some other available way. So Davidson's account needs to be supplanted with this sort of contrastive explanation. And, as talk of "available ways" suggests, *this* explanation may need to appeal to the freedom relation.

Even if we grant that Davidson's account needs to be extended to be properly contrastive, it is not clear that this extension will require appealing to the freedom relation itself. It is true that, when someone acts in a certain way, it is often true that he acts in that way because he believes that some other action was not available. But the reference to *belief* here is crucial. Such an explanation would be equally good whether or not the belief was true or false – that is, whether or not the agent in fact stood in the freedom relation to that action. So even if Davidson's account needs to be extended in something like this way, no appeal to the freedom relation need figure in it. For it is an agent's beliefs about what he is free to do, rather than the freedom relation itself, that figure in this sort of explanation.[76]

So the freedom relation does not figure in the Davidsonian explanation of why agents act as they do, nor does it figure in the contrastive extension of that explanation which also explains why they act in certain ways rather than in other ways.[77] The central point of the

---

[76] Compare Gilbert Harman on moral properties in Harman 1977, pp. 6-9. The broadly internalist assumptions about action explanation on which this sort of argument turns have been called into question in recent years, most powerfully in Williamson 2000, Chapter 3. But even if we grant Williamson's claim that we *sometimes* need to appeal to the truth or falsity of mental states in our explanations (or more precisely to a "factive" mental states like knowledge, which include truth among their necessary conditions) it is not clear how in the present case (or in Harman's case) one goes wrong by only appealing to the beliefs themselves, and not also to that which makes them true or false. And this sort of "local" internalism is enough for the present argument to go through.

[77] A third possibility not canvassed here is the idea that the *absence* of freedom may play a role in explanations of why agents *fail* to act in certain ways. If there is a glass of water on a table, and none of three agents drink it – one because of his hydrophobia, one because he is gagged, and one because he cannot reach the water – the best explanation of why the glass remains full may be that none of the agents was *free to* drink the water. There are a number of questions here. One is whether this is indeed a good explanation, or whether it is merely an informal gloss for a proper explanation, one in which the absence of freedom does not occur. This question is, I think, one worth pursuing; the topic of what we might call *omission explanation* has not been subject to nearly as much study as its better-known partner, namely *action explanation*. Another question is whether and how the fact that a relation's *non-instantiation* figures

arguments for these claims may be put as follows. If a broadly Davidsonian view is correct, then the complete explanation of agents' actions needs to invoke only their *psychological states*. But the freedom relation is not a psychological state, nor is there any psychological state required for explanation such that standing in the freedom relation to an action is a necessary condition for being in that state.[78] Thus the freedom relation is dispensable from the point of view of action explanation.[79]

The claim that the freedom relation does not figure in Davidsonian action explanation constitutes a defense of premise (2), of course, *only if* a Davidsonian view of action explanation is in fact true. As I noted above, I think that most of the purported counterexamples to this view are unsuccessful, and the alternatives to it – notably the agent-causal view – are unsatisfactory. So I think that we should accept the Davidsonian view, provided that it allows us to make room for the aspects of agency of which it makes no mention.

The aspect of agency with which I am concerned here is the freedom relation. I think that we *can* make room for the freedom relation while at the same time accepting a Davidsonian view of action explanation, and so accepting that the freedom relation does not figure in the explanation of why agents act as they do. That is what I will try to do in Chapter 4. Since the foregoing argument for premise (2) rests on the viability of the Davidsonian view of action explanation, and since the viability of the Davidsonian view rests in turn on our being able to give a positive account of the freedom relation that is compatible with it, my defense of the explanatory dispensability of the freedom relation remains, until the next chapter, incomplete.

3.6 A puzzle about ability

Let me briefly restate the argument that the freedom relation does not figure in explanations:

(1) If the freedom relation figures in any explanation at all, then it figures in action explanation
(2) The freedom relation does not figure in action explanation
(C) The freedom relation does not figure in any explanation at all

---

in good explanations (if indeed it does in this case) gives us reason to believe that that relation ever *is* instantiated, which is what is at issue here.

[78] As would be the case, for example, if we took *knowing* that one is free to perform an action to be a psychological state and claimed that such a state needed to be appealed to in action explanation. For the reasons given in the previous footnote, I do not think that this is the case.

[79] Compare the relationship between consciousness and psychological states in Chalmers 1996. Just as I have argued that freedom is not needed for action explanation, since such explanation needs to invoke only psychological states, so Chalmers argues that (modulo some open questions about the relationship between consciousness and causation) *consciousness* is not needed for action explanation, since such explanation needs to invoke only psychological states (pp. 11-16).

In the previous section I mounted a case – as I admitted there, a yet inconclusive one – for premise (2). In this section I want to consider what seems to be the strongest objection to premise (1).

For the freedom relation to figure in explanations, it need not figure *explicitly* in explanations. If there were some relation that indubitably figured in explanations, and if it *turned out* that standing in the freedom relation were a necessary condition for standing in that relation, then it would turn out that the freedom relation did figure – implicitly – in explanation, after all. For example, *acceleration* seems to be a relation (between an object and a rate of change of velocity) that figures indispensably in explanations. If it turned out that standing in the freedom relation was a necessary condition of standing in the acceleration relation, then it would turn out that freedom was indispensable to explanation after all. This case, of course, is not plausible. In the remainder of this section I want to consider a case that is somewhat more plausible.

In Chapter 2.7, I argued that, while the freedom relation is not reducible to the ability relation, the converse may be true. In particular, I conjectured that the correct analysis of the ability relation may be:

(ABLE) S has the ability to A just in case S is normally, in virtue of his intrinsic properties, free to A

Yet, conjoined with the claims advanced in this chapter, (ABLE) seems to create a puzzle. Let me explain.

It seems very plausible that *abilities* have a role to play in explanations. Consider, for example, the conditional fact that Roger Federer would defeat me if we were to play tennis. This seems to be true, and the best explanation of this fact would seem to invoke various abilities that Federer has that I do not: the ability to hit a very fast and accurate serve, for example. So the ability relation, unlike the freedom relation, seems to have a role to play in explanations.

But here is our puzzle, which may be stated as a trilemma:

(1) The freedom relation does not figure (explicitly or implicitly) in explanations
(2) The ability relation does figure in explanations
(3) The ability relation is reducible to the freedom relation (as (ABLE) claims)

The conjunction of (1), (2), and (3) cannot be true. Which should we reject?

We should not reject (1). This claim is, I think, a cornerstone of the proper account of the freedom relation. I have tried to make the case for this claim here, and will continue the case in the next chapter, where I will sketch a positive account of the epistemology of the freedom relation that does not rely on freedom having an explanatory role to play. So rather than taking this puzzle as an objection to the main argument of this chapter, I take

it as revealing that one of two intuitive claims about the ability relation – (2) or (3) – is false.

Which of these claims we reject is, for the purposes of this essay, a matter of indifference. Perhaps since it is (3) that is the more novel claim, we should reject it. Perhaps (ABLE) fails because abilities figure in explanations in a way that mere generalizations of the freedom do not. Since (ABLE) is for present purposes merely a conjecture, this rejection of it is compatible with the arguments of this essay. Or it may be that it is (2) we should reject; despite its intuitive appeal, it is not obvious that abilities are fit to play the explanatory role that (2) assigns to them.[80] But in neither case does this puzzle gives us reason to reject (1).

3.7 Freedom and immediate justification

I take the foregoing arguments to have provisionally established a simple conclusion: there is no broadly inferential justification, either from enumerative induction or from inference to the best explanation, for our beliefs about freedom. To accept this conclusion is not yet to accept skepticism about freedom, for all that has been established is a conditional claim: if we have justification for our beliefs about freedom, then that justification is not inferential.

We should not accept skepticism about freedom, for the antecedent of this conditional is, I think, satisfied. In the next chapter, I will argue that we have justification for our beliefs about freedom, but not because we are justified in *inferring* claims about freedom from other claims which we are already justified in believing. Our justification for believing claims about freedom is non-inferential or, as I will say, *immediate*.

There are some who argue that immediate justification in this sense is not possible: that all justification involves either explicit or implicit inference. I do not accept these arguments, but I will not offer anything like a general defense of the possibility of immediate justification here.[81] My aims are more modest. I shall assume what I think is plausible, namely that *perception* sometimes provides us with immediate justification. For example, I now hear music playing in the other room, and this immediately gives me justification for believing that there is music playing in the other room.

In the next chapter, I will make my central positive argument. This is that our faculties of bodily awareness represent the freedom relation, and that this representation is perceptual. In short, the freedom relation is something that we normally perceive. (If one insists on reserving the term "perception" for the external senses, then I am happy to call our awareness of the freedom relation *quasi-perceptual*, provided that this is not taken to suggest any sort of epistemic deficiency). Coupled with the claim that perception can provide immediate justification, this account will yield an explanation of how we have *immediate justification* for the instantiation of the freedom relation.

---

[80] Compare the problems about assigning an explanatory role to *dispositions* raised in Prior, Pargetter, and Jackson 1982.

[81] See Pryor 2005, which sketches a compelling case against the denial of immediate justification.

If this account is successful, then it will give us sufficient grounds to reject *one* kind of skepticism about freedom, the one which has been at issue in this chapter: a skepticism which is willing to accept something like our ordinary ways of gaining evidence about the world.[82] The argument against this skeptic will proceed by way of parallel: insofar as external sense perception is a way of gaining immediate justification about our environment, so is bodily awareness a way of gaining immediate justification about what we are free to do. There are also some epistemic questions that are specific to the case of freedom. These I will address in Chapter 5.

3.8 The does-can gap

Before proceeding to my own positive account, it may be helpful to summarize the conclusions of the previous three chapters. In Chapter 1, I identified the freedom relation. In Chapter 2, I argued that this relation cannot be *reduced to* any relation that does not include the freedom relation among its constituents. In this chapter, I have argued that claims about the instantiation of the freedom relation cannot be *inferred from* claims about any distinct set of facts.

The previous three chapters have thereby established what we might call a *does-can* gap between facts about freedom and facts that are not themselves about freedom. This gap is *ontic* (in light of Chapter 2) and *epistemic* (in light of Chapter 3). In these respects, it is parallel to the more familiar *is-ought* gap. As facts about freedom cannot be reduced to or inferred from facts that are not themselves facts about freedom, so – according to those who claim there is an is-ought gap – can facts about normativity neither be reduced to nor inferred from facts that are not themselves facts about normativity.[83]

It should not be surprising, I think, that there is this sort of parallel between freedom and normativity. Both are aspects of the world that seem indispensable from the first-personal point of view of an agent, yet which seem difficult to reconcile with the world as we know it from the third-personal perspective of the sciences. Facts about propensities and probabilities, it seems, cannot *add up* to facts about freedom, just as facts about

---

[82] I consider a more radical sort of skepticism about freedom in the Conclusion.

[83] The canonical expression of the view that there is such a gap is David Hume's:

> I am surprised to find, that instead of the usual copulations of propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not. This change is imperceptible; but is however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, it is necessary that it should be observed and explained; and at the same time that a reason should be given; for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it.

Hume 1739; Book 3, Part 1.1. It does no injustice to Hume's point, I think, to put it in precisely the terms in which I have made my point about the freedom relation. In these terms, Hume's claim is just that the *recommendation relation*, which is a relation between agents and acts that is normally expressed by the modal auxiliary "ought," is neither reducible to nor inferable from any set of facts that do not make reference to the recommendation relation itself.

desires and commands cannot *add up* to facts about normativity. So I think that the parallel between freedom and normativity here runs quite deep.

One of the advantages of underscoring this parallel is that it opens up a range of more substantive paths of inquiry. It is worth asking to what degree the multitude of views developed in contemporary approaches to the is-ought gap may also be applied to the question of freedom, where there seem to be a rather more limited range of views available. For present purposes it is the varieties of *realism* on offer in contemporary metaethics that are most salient. In particular, the positive view that I will develop in the remainder of this essay is very close in spirit to *intuitionism* about the normative, on which normative claims are a distinct and irreducible set of claims for which we have immediate justification.[84] The view I will develop is actually in at least one way better off than normative intuitionism, for rather than needing to appeal to a distinct faculty of "intuition," my view appeals only to the deliverances of the mundane and familiar faculties of bodily awareness.[85]

So I think that making explicit the parallel between these "gaps," as I have, makes clear to us the *range* of views of freedom that there are to be developed. This range includes what I think is the *true* view, which I will set out in the remainder of this essay.

[84] This sort of view has a complicated history which I will not try to parse here. See Huemer 2005 for a recent defense.

[85] In this respect my view is closer to a similar view of color, which we might think of as responding to the *chromatic-monochromatic* gap. On this view, colors are simple and irreducible properties of objects, ones which are revealed to us in a single sense modality, namely vision. See Campbell 1993.

The Possibility of Freedom

Chapter Four: The Perceptual Thesis

4.1 The experience of freedom

In this chapter I will propose an account that will solve the epistemological problem raised in the previous chapter, and which is the foundation of an adequate theory of the freedom relation. This account will not be especially novel. Indeed, I think that the main contribution of this chapter is simply to state and defend as carefully as possible an idea that is, in its broad outline, extremely familiar. Let me say what I take that idea to be.

The idea, most concisely stated, is that freedom is something we *experience*. Part of *what it is like* to be an agent is to experience oneself as being free to perform various actions. Any philosophical *theory* which purports to give a comprehensive account of agency is obliged to take note of this experience, even if only to dismiss it as an illusion. This is the idea that will be the linchpin of the positive account that I will propose in this chapter.

There are a variety of phenomena that might be denoted by the phrase "the experience of freedom." In the next section I will explain precisely what I mean when I speak of such an experience. In this section I want to set to one side some phenomena that are sometimes denoted by this phrase but with which I will not be concerned here.

One thing that might be meant by "the experience of freedom" is what might be more accurately called "the experience of *acting*." It is sometimes claimed that when we act, we have an experience as of bringing something about, rather than merely of having something *happen* to us. I find it very plausible that there is such an experience, and that it may have some evidential value of its own.[86] But it will not be my concern here. Recall that the freedom relation is essentially a relation to acts we do *not* perform. Since the experience of acting, if such there be, is an experience of acts that we *do* perform, it is not immediately relevant to giving an account of the freedom relation. When I speak of "the experience of freedom," I will be referring to an experience that is in the first place of acts that we do *not* perform.

Another thing that might be meant by "the experience of freedom" is what might be more accurately called "the *conviction* of freedom." There may be some philosophical beliefs that are widely, strongly, and "pre-theoretically" held. *Perhaps* the belief that we are free to perform actions that we do not perform is among these, and perhaps this belief is so basic that it deserves the name of "experience." It is an interesting question whether we in fact do have such a belief, and whether such beliefs, in this case and more generally, ever ought to revised in the light of philosophical argumentation.[87] But this belief, if

---

[86] A useful recent discussion of this sort of experience is Siegel 2005.

[87] On the more general case, see Kelly 2005. An interesting discussion specific to the case of freedom occurs in van Inwagen 1983, pp. 153-161. Distinguish this position, which essentially makes a *psychological* claim, from the *epistemic* position according to which agents always have "unearned"

such there be, is not what I am referring to when I speak of "the experience of freedom." I will be using the phrase "experience" in a more literal sense, to refer to a *presentation* as of things being a certain way. In this sense of "experience," one may in principle have the experience of freedom while lacking the belief that one is free, and conversely.

So when I speak of "the experience of freedom," I will be referring to neither an experience of acting nor of a belief in freedom. Instead, I will be using the phrase in what I think is something like its most natural sense. I will be referring to an experiential state immediately representing the agent who bears it as being free to perform various actions. That formulation itself remains a bit obscure; in the next section I will try to illuminate it.

4.2 The proposal stated

Let me state the positive proposal, and then proceed to explicate it. According to *the perceptual thesis*:

(PT) (i) For a typical agent S, it is normally part of the representational content of S's bodily awareness that S stands in the freedom relation to some actions, *and*

(ii) This representation is normally genuinely perceptual.

This formulation includes several qualifications, and several terms of art. Let me begin with the qualifications.

The first qualification restricts the domain of agents to whom (PT) applies. There are I think actual agents, even ones with sophisticated cognitive capacities, to whom (PT) does not apply. And it is for all I say here *possible* that there are creatures very much like us but for the fact that (PT) does not apply to them. But I say that for most typical agents, with myself and others like me standing as canonical examples, (PT) is true.

The other qualifications, namely the "normally" that occurs in both (i) and (ii), serve to *strengthen* (PT). It is not merely an occasional occurrence that agents perceptually represent the freedom relation, and (PT) would not be true if it were. (PT) asserts that it is normally the case that agents perceptually represent the freedom relation – this is a situation that tolerates exceptions but otherwise normally obtains.

The crucial terms of art that figure in (PT) are *bodily awareness*, *representational content*, *the freedom relation*, and *perceptual*. Let me explain what I mean by each of these.

By "bodily awareness" I refer to an agent's inner sense of the properties of his body, including for example the relative position of his limbs, the pressure being applied by and upon him, and his sense of balance. This list of the contents of bodily awareness is not

---

justification for the belief that they are free to perform actions they do not perform; this latter position I will discuss below in Section 7.

all-inclusive. Indeed, what is most crucial here is what is *excluded*. Bodily awareness excludes both the awareness of the properties of his body that an agent acquires through his external senses (including the sense of touch) *and* his introspective awareness of his own mental states. As for bodily awareness itself, I mean to leave as open as possible many of the empirical questions in this neighborhood, such as whether there is one faculty of bodily awareness or, as seems more likely, several distinct faculties working in tandem. So I take myself simply to be *identifying* the sensory modality through which we experience freedom, while remaining as neutral as possible on the workings of the faculty or faculties that undergird it.

In speaking of the *representational content* of an experience, I mean only the information conveyed by that experience. This is of course an elucidation rather than a definition, for the notion of an experience conveying information is itself open to a variety interpretations. There are several theories available here, and I shall try to remain as ecumenical as possible with regard to these theories in what follows. The only constraint I will impose on such theories is that they allow for the sorts of experiences described in (PT). So I will be speaking throughout of *having an experience as of* standing in the freedom relation, and of *perceiving that* one stands in the freedom relation. *Any* theory of the contents of experience that can make sense of these phrases will be adequate for my purposes.

The freedom relation is something I have already discussed at length. To say that the freedom relation figures in the contents of experience is just to say that this very relation, the one which I have claimed is neither reducible to nor inferable from any other facts, is something that normally figures in the contents of an agent's bodily awareness.

Let me also say something about the *relata* of the freedom relation as it figures in bodily awareness. The *subject* of the freedom relation is, plausibly, the agent himself: the one who is the bearer of the experience in question. The *object* of the freedom relation is normally an action. But it is not plausible that the freedom to perform *any* action whatsoever can figure in the contents of experience. I may, for example, now stand in the freedom relation to robbing the First National Bank, but this is not something that plausibly figures, in its entirety, in the contents of my bodily awareness. I think that it is plausible that the objects of the freedom relation here are in the first place *basic actions*, which I take to at least *include* bare movements of one's body, such as moving one's arm in a certain manner.[88] These are the sorts of actions that fall within the domain of the freedom relation, when it figures in the content of bodily awareness.

The final term of art that figures in (PT) is the one that occurs in clause (ii): perceptual. The demand imposed by clause (ii) is that the experience of freedom is not *mere* experience. It rather has an important *epistemic* status. The experience of freedom, because it is perceptual, normally gives its subject justification for believing that he *is*

[88] See Danto 1963 and 1965. There is much more to be said on this point. Ultimately, I suspect that the statement given here may be in a certain way circular, for a basic action may *just be* any action that one is immediately free to perform; that is, we can give a definition of basic action only by invoking the freedom relation itself.

free to perform various actions. This is the sense in which (PT), if true, yields a solution to the epistemological problem raised in the previous chapter. Below, I will have more to say below about the demands of perception, and whether in the case of the experience of freedom these demands are in fact met.[89]

4.3 Four objections to (PT)

There are a number of ways one might object to this proposal. It may be helpful to divide these into two types, under each of which two more particular objections fall:

Objections to the possibility of (PT):
- (P1) The freedom relation *cannot* be represented in experience
- (P2) The freedom relation *cannot* be perceived

Objections to the truth of (PT):
- (T1) The freedom relation *is not* represented in experience
- (T2) The freedom relation *is not* perceived

Since I hold that (PT) is true and therefore possible, I am obliged to answer all four of these objections. Answering these objections will, if successful, also advance the case for my own positive view. For so far I have simply *stated* (PT). In the course of responding to these objections, I will offer my own positive *arguments* in favor of the truth of (PT). Here let me briefly sketch the general nature of each of these objections.

According to (P1), the freedom relation is not a possible content of experience. In developing this objection, it is crucial not to assume any especially contentious view of experience, one which the defender of (PT) may reasonably deny. For example, some hold that the *object* of experience is something like a two-dimensional array of colored patches. If this is true, then the representational capabilities of experience are severely limited. Experience could not represent freedom, but many other seeming contents of experience would also be excluded. I will argue that it is difficult to defend (P1) once we have renounced this sort of strategy.

According to (P2), the freedom relation is not a possible object of *perception*. Perception is to be distinguished from representation, as I have already said, primarily on *epistemic* grounds. To claim that freedom is an object of perception is not to claim merely that it is something that is represented by agents.[90] It is also to claim that these representations give them justification for that which it represents. Again, it will be important for the proponent of this objection not to impose any tendentious requirement on our ways of gathering evidence through experience. I will develop a version of this objection that does not rely on any such requirement.

---

[89] In Sections 4 and 9.

[90] Though it is to claim *at least* that, on the assumption that perception requires representation. This assumption is perhaps questionable. Assuming it does not however make things any *easier* for me, since this assumption effectively makes the perceptual relation *more demanding* than it might actually be. I will return to this assumption below in Section 7.

The defender of (T1) is willing to grant that freedom *could be* represented in experience, but denies that it is represented in experience. It is difficult to assess this sort of argument because we lack any reliable way of determining the contents of experience, but for nonce appeals to introspective intuition. I will offer an argument *against* (T1) that does not depend on any claim about what I find when I look within.

The defender of (T2) is willing to grant that freedom *could be* perceived, but denies that it is perceived. This objection is I think the most difficult one to answer, because the demands on the perceptual relation are such that *a priori* argumentation alone does not suffice to determine whether they are met. Rather than giving anything like a demonstration that the freedom relation is perceived, I will answer this objection by saying, first, what demands must in fact be met for the freedom relation to be something that is perceived and, second, that the evidence that we have indicates that these demands are in fact met at least as well as they are in paradigm cases of perception, such as our perception of objects in our immediate environments.

It should be clear that these objections stand in a sort of dialectical hierarchy. If (P1) is true, then so is (T1); similarly for (P2) and (T2). And if we assume that something can be perceived only if it is represented by the perceiver, then (P1) entails (P2), and (T1) entails (T2). I will take these objections in this order, proceeding from the most sweeping objection, (P1), to the most particular, (T2).

4.4 Freedom can be represented in experience

Views like mine are sometimes rejected on the grounds that they make something like a category mistake from the outset. Here is J.S. Mill against a view of this kind proposed by William Hamilton:

> Consciousness tells me what I do or feel. But what I am *able* to do, is not a subject of consciousness. Consciousness is not prophetic; we are conscious of what is, not of what will or can be.[91]

It would indeed be devastating to the present view if something like this claim were true. But I see no reason to accept it.

As noted in the previous section, it is true that certain views of the *objects* of experience place severe limits on the *contents* of experience. If these objects are something like *sense-data*, then perhaps freedom could not be represented by them. But we have no reason I think to accept so severe a limitation on the objects of experience. Absent an *argument* that a particular sort of property of objects cannot be represented in experience, it would be unwarranted to assume that the contents of experience are any less various than the objects which those experiences represent. Indeed, if we think of experience as being in a certain way "transparent," then our presumption should be that any property

[91] Mill 1889, p. 580

that might be a property of objects might in turn be represented in experience.[92] There is no special problem of determining how those properties could be "constructed" out of the impoverished materials of experience, for we lack any prior ground for the assumption that experience is so impoverished.

There a way of reading this quotation from Mill, however, on which he is making a rather different point. On this reading, Mill does not straightaway deny that freedom could be *represented* in experience; rather he denies that we could *find out* about freedom through such representations alone, that consciousness could "tell me" about such things. Indeed, as Mill's remark proceeds, it becomes clear that his question is not what experience can represent, but what we can (or cannot) come to *know* on the basis of experience:

> We never know that we are able to do a thing, except from having done it, or something equal and similar to it. We should not know that we were capable of action at all, if we had never acted.[93]

This may or may not be true. But in any case, it is not a psychological point about the possible contents of experience, but an *epistemic* point about what we can come to know on the basis of experience.

In short, on the most plausible reading, Mill is not claiming that freedom cannot be represented in experience, but rather that freedom cannot be *perceived*. Let us turn then to that objection.

4.5 Freedom can be perceived

What argument might there be that freedom cannot be perceived? One way of arguing this point is by analogy. We compare freedom to other properties that, intuitively, are not the sort of thing that can be perceived. Unless we can find some point of disanalogy between these properties and freedom, then we ought to grant that freedom cannot be perceived either. Here I will consider what seems to me the strongest such argument, one suggested by Hume: this argument turns on an analogy between freedom and *dispositions*.

It will be useful to begin by distinguishing two ways in which some property P might be perceived. Say that a property P is *mediately perceived* just in case there is some perceptible property Q, distinct from P, which is known on the basis of broadly inductive reasoning to be *reliably correlated* with P. This is the sense in which a driver perceives the level of gas in his tank: he perceives some *other* property, namely the height of the gas gauge, which he knows to be reliably correlated with the level of gas in his tank. Say that a property P is *immediately perceived* just in case it is itself the perceptible property. This is the sense in which the driver perceives the height of his gas gauge. There is no distinct property which the agent knows to be correlated with the height of his gas gauge. Rather, the height of his gas gauge is *itself* an object of perception for him.

---

[92] See Harman 1990.
[93] Mill 1889, p. 580.

The Humean objection that I will consider in this section denies that freedom can be, in this sense, immediately perceived. The Humean is willing to grant that freedom may be *mediately* perceived: that is, that we might on the basis of inductive reasoning know it to be correlated with some perceptible property. But he denies that it is the sort of thing that can be immediately perceived. This objection, if sound, would be devastating to the main argument of this essay. For in the previous chapter I argued that we never have any inductive evidence that the freedom relation is instantiated. *A fortiori*, we never have any inductive evidence that the freedom relation is instantiated whenever some property distinct from it is instantiated. So by the arguments of the previous chapter, we have justification for believing that the freedom relation is instantiated *only if* it is something that we sometimes immediately perceive.[94]

What is the Humean argument? In the *Enquiry* Hume makes the following claim:

> This proposition, that causes and effects are discoverable, not by reason but by experience, will readily be admitted with regard to such objects, as we remember to have once been altogether unknown to us; since we must be conscious of the utter inability, which we then lay under, of foretelling what would arise from them. Present two smooth pieces of marble to a man who has no tincture of natural philosophy; he will never discover that they will adhere together in such a manner as to require great force to separate them in a direct line, while they make so small a resistance to a lateral pressure.[95]

The letter of Hume's argument is against the possibility of discovering facts about the causal relation in certain ways, but a precisely parallel argument tells against the possibility of discovering facts about the *disposition relation* in those ways. For consider the disposition that these two pieces of marble have to separate when lateral pressure is applied to them. Is this disposition the sort of thing that could be immediately perceived in them, by any of the senses? Intuitively not: one must apply the relevant stimulus conditions to the pieces of marble to discover whether or not they have this property. So the disposition of these thing is at best *mediately* perceived: one might perceive that the pieces of have some perceptible property, say smoothness, which one knows to be correlated with the disposition to separate when lateral pressure is applied. Or, in the case Hume is considering, where the objects are "altogether unknown" to the observer, their dispositional properties are not perceived at all.

It will be clear that this point generalizes. The most general statement of this view we might call *hiddenness of dispositions*:

---

[94] This is not to say that freedom could not be mediately perceived. Indeed, I think cases of this are widespread: we perceive the freedom of others, for example, only mediately. The point here is rather that if the freedom relation is never immediately perceived, then we never have *any* justification for believing that it obtains, and so it is never mediately perceived either.

[95] Hume 1748, Section 4, Part 1.

(HD) For any object x, stimulus C, and response R, it is impossible that it is immediately perceptible that x, C, and R stand in the disposition relation

For precisely the sort of reasons that Hume gives, (HD) seems a very plausible thesis about dispositions.[96] And this gives rise to the argument by analogy that the freedom relation cannot be immediately perceived either. The disposition relation and the freedom relation are both concerned, intuitively, not with what in fact happens but with what *might* happen. We should therefore conclude that the freedom relation is also "hidden" in this respect: it is not the sort of thing that can be immediately perceived.

This argument trades on an analogy between dispositions and freedom that I think we should reject, for the analogy is mistaken in at least two regards. One disanalogy is epistemic; this point is indecisive, I admit, against the Humean. The second disanalogy is metaphysical: it turns on a basic difference between the nature of the freedom relation and that of the disposition relation. This second disanalogy is, I think, decisive against the Humean argument, and so against inferring the denial of (PT) from the truth of (HD)

The first, epistemic, disanalogy is the following. The Humean argument assumes epistemic parity between the information we can gain about external objects on the basis of sense perception and the information we can gain about ourselves through bodily awareness. But we might reasonably wonder whether this sort of epistemic parity obtains: it could be that there are some properties that are only perceptible through bodily awareness. So it is at least possible that (HD) is true, that the freedom relation is importantly analogous to dispositions, and yet that (PT) is also true.

But this disanalogy is, as I said, at best indecisive against the Humean. For we have given no positive reason to think that bodily awareness is more liberal in its deliverances than sense perception, as the development of this disanalogy would require. Indeed, this disanalogy is rather against the spirit of my arguments thus far, which have turned on the parallels between sense perception and bodily awareness. So this disanalogy seems unpromising. It is also, I think, unnecessary, for there is a different and better way of resisting the Humean argument. This way does not involve postulating an especially permissive epistemology. It rather rejects the metaphysical analogy between the disposition relation and the freedom relation on which the Humean argument turns.

Let us first ask: *why* does the Humean argument succeed in the case of dispositions? A short but informative answer is that the disposition relation is essentially a three-place relation: a relation between an object, a stimulus, and a response.[97] To determine

---

[96] The closest thing I know of to a denial of (HD) occurs in the Gibsonian literature on our perception of the "affordances" of things. See Gibson 1979, Chapter 8. Interestingly, an affordance is not exactly a disposition; it is rather a more complicated relation between an observer and what he *can do* with that object. For example, a bridge might seem to me to have the affordance of being something I can walk upon. This makes tempting the following conjecture: to perceive an affordance is not to perceive a disposition, but rather to conjoin one's inductive knowledge of the dispositions of an object with one's immediate perception of one's freedom with respect to that object.

[97] The failure of the conditional analysis of dispositions lies *not* in taking the disposition relation to be a stimulus-response relation but rather offering too simple an analysis of that relation, namely an analysis in

whether an object has that disposition, it is normally necessary at least to *apply* the relevant stimulus. This is why immediate perception alone does not suffice to determine whether or not some object has a given disposition. In the case that Hume considers, to learn how pieces of marble would behave when pressure is applied to them, one must actually apply the relevant pressure to them. Such a test will not be a sure-fire way of learning of their dispositions,[98] but the application of the relevant stimuli in a range of different scenarios is normally both necessary and sufficient for determining the dispositions of the thing in question.

One of the lessons of Chapter 2, however, was that the freedom relation was importantly different from the disposition relation in this respect. There is nothing that stands as a "stimulus" to the freedom relation, and attempts to have some mental state (such as choice or trying) fill that role lead to erroneous analyses. In the terms in which these issues were once debated, we might say that the freedom relation is *categorical* rather than *hypothetical*.[99] To be free to perform some action is not for it to be the case that one would act in a certain way in response to certain stimuli, but simply for one to stand in the freedom relation – which is, unlike the disposition relation, an essentially two-place relation – to that action.

In short, I think the appeal of the Humean argument for (HD) arises from the fact that the disposition relation is a hypothetical one. Since the freedom relation is not hypothetical, this argument gives us no reason to accept a similar thesis about freedom. For the aspect of dispositions that gives rise to the Humean argument is not also an aspect of freedom.

It would be nice to also have a *positive* analogy here: to point to another property that is, like the disposition relation, concerned with how things *might* behave but which is also immediately perceptible, as I claim that the freedom relation is. I do not know of any such relation: in this sense the epistemology of the freedom relation is in a certain way unique.[100] But to make the case that the freedom relation is at least a possible object of

---

terms of a counterfactual obtaining between the stimulus and response. As I suggested in Chapter 2.7, the disposition relation should rather be thought of as a certain sort of *generalization* of the stimulus-response relation.

This proposal also seems to fall afoul of alleged cases of dispositions that do not have stimulus conditions (for example in Manley and Wasserman 2008). I am inclined to think that these cases are something of a grab-bag, and that none of them are dispositions that do not have stimulus conditions. Some are dispositions whose stimulus conditions we normally do not mention, while others genuinely lack stimulus conditions but are not dispositions at all – that is, they are of a different ontological *kind* than fragility, solubility, and so forth. But an adequate defense of this proposal about dispositions is beyond the scope of this essay.

[98] As it would be if the conditional analysis of dispositions were true.

[99] See Austin 1956.

[100] Even when we consider cases of other categorical relations that are also concerned with how things *might* behave, their epistemology is quite different from that of the freedom relation. Consider *chance*. Like the freedom relation, the chance relation is categorical rather than hypothetical. But unlike freedom, chance is not plausibly a possible object of immediate perception. So I have not argued that *any* categorical relation is immediately perceptible. Rather I have argued that since the Humean argument that dispositions are not immediately perceptible turned on the fact that dispositions are hypothetical, it does not also apply to freedom. The epistemology of categorical relations needs to be made on a case-by-case basis, with some

perception, it suffices to draw a principled distinction between it and relations which are *not* possible objects of immediate perception, notably the disposition relation. That is what I have tried to do in this section.

4.6 From possibility to actuality

The objections to which I responded in the previous two sections were claims about impossibility. The claim that it is *impossible* for a certain psychological or epistemological relation to be instantiated is a very strong one. My responses to these objections, in turn, needed to show very little. I argued that claims of the impossibility of (PT) rested on tendentious restrictions on the possible contents of experience, on the one hand, or faulty analogies to other relations, on the other. These arguments made at least a preliminary case, I think, against the impossibility of (PT).

But I am claiming that (PT) is not merely possible but *true*, and this is a more difficult claim to establish. Our problem is redoubled because the claim that we are interested in, (PT), is not one that admits of straightforward confirmation or disconfirmation. For it includes claims about the contents of experience which seem to be available only through introspection (and perhaps not *even* through introspection) . Furthermore, the *descriptive* claim made by (PT) is bound up in various ways with the *normative* claim that we are *justified* in believing claims about freedom. In short, determining the *truth* of (PT), rather than its mere possibility, is not a straightforward matter.

In the next three sections I will try to overcome these difficulties and present a case that (PT) is true. This case will not be airtight. In the case of representation, my arguments will turn on a contentious principle about how to determine the contents of experience. In the case of perception, I will argue that the truth of (PT) cannot even be decided by *a priori* methods alone, though I will argue that the extant evidence weighs in favor of (PT). So my arguments here are not *decisive* ones, in the sense of removing any doubt that (PT) is true. Nonetheless they are, I think, *good* ones.

4.7 Freedom is represented in experience

My argument that freedom is represented in experience will proceed in two steps. In this section, I will defend a conditional claim about the relationship between our justification for believing claims about freedom and the representation of freedom in experience. In the next section, I will defend a quite general principle about the contents of experience. Taken together, this claim and this principle entail that freedom is represented in experience.

The conditional claim is the following:

(JE) The belief that the freedom relation is instantiated is justified *only if* the freedom relation is represented in experience

---

of these, like freedom, being immediately perceived, while others, like chance, being known on other grounds.

I accept (JE) because I find the following argument to be sound:

(1) The belief that the freedom relation is instantiated is justified *only if* it is either inductively justified *or* it is justified by immediate perception
(2) This belief is not justified inductively
(3) If this belief is justified, then it is justified by immediate perception (by 1 and 2)
(4) This belief is justified by immediate perception *only if* the freedom relation is represented in experience
(JE) The belief that the freedom relation is instantiated is justified *only if* the freedom relation is represented in experience (by 3 and 4)

Since this argument is clearly valid, a defense of (JE) requires only a defense of the truth of each of its premises.

Premise (3) is true if premises (1) and (2) are. And the previous chapter was an extended defense of premise (2). So a complete defense of this argument requires only a defense of premise (1), and of premise (4).

Let me begin with premise (4). This premise turns on the assumption that the perceptual relation is instantiated only if the representation relation is instantiated. This is an assumption that I briefly noted above and, as I said there, I grant that it may be questionable. A proper defense of this assumption would take us far afield, into questions about the nature of experience that I do not have space to adequately address here. This assumption may be thought of as something like a working (and, I think, plausible) account of the relationship between experience and epistemology. If this assumption turns out to be false, then the *letter* of (PT) would have to be altered, though I suspect that something like its spirit might still be true. In any case, the relationship between experience and epistemology would be more *complicated* than it is on this account. Of course, the truth may be complicated. My best defense of this assumption is that it is a widely-held and plausible position and, more importantly, that assuming it seems the most effective way to adjudicate the truth of a perceptual account of freedom without delving into deeper questions about the nature of experience.[101]

If we accept premise (4), then the truth of (JE) turns on the truth of premise (1), according to which the belief that the freedom relation is instantiated is justified only if it justified inductively or if it is justified by immediate perception. Since these options are not exhaustive, a defense of premise (1) requires ruling out other ways in which the belief that the freedom relation is instantiated might be justified. The salient omissions here are ways of gaining justification about the freedom relation that do not rely on anything like empirical evidence. Let me review two such ways, one familiar and one less familiar.

The familiar way proceeds by way of moral responsibility. According to *the argument from moral responsibility* (AMR), it is *a priori* that normal agents are morally responsible

[101] A useful recent criticism of this assumption, which also includes extensive testimony to its orthodoxy, is Alston 2005.

for what they do. And, according to this argument, it is also *a priori* that an agent is morally responsible for he does *only if* he stands in the freedom relation to some actions. Therefore it is *a priori* that any normal agent stands in the freedom relation to some actions.[102]

This argument is an old one, and I cannot do full justice to it here. Let me briefly note two ways in which one might resist (AMR). One is by being a *skeptic* about moral responsibility, and therefore denying that we have justification for believing the first premise. The other is by being what is sometimes called a *semi-compatibilist* about moral responsibility, and therefore denying that we have justification for believing the second premise. Both positions have advocates, and both are motivated on independent grounds.[103] There is no quick way of resolving these questions, and so no quick way of accepting or rejecting (AMR).

The relationship between this argument and the one I am advancing here is as follows. If one rejects (AMR), then one should therefore reject *this* way of denying (2) (there remains another way, which I will consider presently). If one accepts (AMR), then while the epistemology of the freedom relation offered here may yet be correct, one will not find *this* argument for it convincing. But, as I said, I will not try to evaluate (AMR) here. Indeed, one of the hopes of this chapter is to develop an epistemology of the freedom relation that is *independent* of contentious questions about moral responsibility.

Another way of rejecting this premise does not claim that we have non-empirical justification that *entails* that the freedom relation instantiated. It rather claims that we have non-empirical justification for the instantiation of the freedom relation itself. This view has not to my knowledge been developed in print, but it follows naturally from an idea of Crispin Wright's. Suggests Wright:

> Suppose there is a type of rational warrant which one does not have to do any specific evidential work to earn; better, a type of rational warrant whose possession does not require the existence of evidence – in the broadest sense, encompassing both *a priori* and empirical considerations – for the truth of the warranted proposition. Call it *entitlement*.[104]

Wright develops this idea with regard to our justification for believing that there is an external world. But the application of this idea is potentially broader than that. For our belief that we are free to perform actions that we do not perform is perhaps like our belief that there is an external world in seeming both as well-justified as any belief could be *and* in seeming to be the sort of belief for which there could be no *evidence*, however broadly evidence is construed.

---

[102] This is the central positive argument of van Inwagen 1983; see especially pp. 161-162.
[103] For a recent discussion of the first, see Rosen 2004. The second premise has gotten much more attention in recent decades, largely due to the fact that Frankfurt 1969 seems to offer *counterexamples* to this allegedly *a priori* claim. For a recent defense of this case against the second premise of (AMR), see Fischer 2002.
[104] Wright 2004, pp. 8-9.

Of course, I deny at least the second part of this claim: there is, according to (PT), ample evidence that the freedom relation is instantiated. Nonetheless, I think this account is an intriguing and undeveloped one, and more needs to be said about it than I will say here. The crucial point to make is that we can do better: (PT) provides an account of the epistemology of freedom on which our beliefs are justified *and* we have evidence for them. Nonetheless, for someone who wants to evade the present argument, this way of doing so remains, for all I have said, available. Indeed, it seems to me the *most plausible* way of denying the epistemological argument for (PT).[105]

With the caveats indicated by the foregoing discussion in place, let us proceed on the assumption that (JE) is true. The question then is: how do we get from the *conditional* claim expressed by (JE) to the claim that freedom *is* in fact represented in experience? Answering this question will be the topic of the next section.

4.8 Freedom is represented in experience, continued

Let us take a step back. Recall that the view that I am advocating here, namely (PT), is not an especially novel one. Attempts to give a broadly "introspectionist" defense of freedom have been around for centuries.[106] But such accounts always seem subject to the objection that their proponents seem to be drawing very strong epistemological consequences from the subjective and variable evidence of what they find when they look within.

As will become clear in this section, my account proceeds in precisely the opposite direction. I will proceed *from* epistemology *to* the contents of experience, rather than vice-versa. This procession may be subject to various objections, but it is *not* subject to the objection that it relies too heavily on introspection. For this account grants very little weight at all to introspection. It does, however, grant heavy weight to our beliefs and to our antecedent confidence that our beliefs are justified.

My argument will turn on a quite general principle about the contents of experience, which I will call the *Permissive Principle*:

(PP) *If* (i) someone is antecedently confident that his belief that p is justified, *and*
(ii) his belief that p is justified only if some relation figures in the content of his experience, *and*
then he ought to be confident that R figures in the content of his experience.

This principle is so-called because, when there is doubt about the contents of experience, it assumes a more permissive construal of those contents.

---

[105] Something similar is true I think in the debates over external world skepticism, where Wright's position seems the most plausible alternative to the "dogmatism" of Pryor 2000. See White 2006 for a recent defense of the preferability of Wright's position on this front.

[106] Reid 1788, Essay 4, Chapter 6, is a classical statement of this sort of view. See Lehrer 1960 for a somewhat more recent defense of these sorts of views.

The permissiveness of (PP) is, I think, warranted. It is warranted by the following two claims. First, that our only immediate access to the contents of experience is introspection. Second, that introspection is sometimes silent or uncertain on what exactly the contents of experience are. According to (PP), when we face this sort of uncertainty we should determine the contents of the experience in the way that best accords with our confidence in our antecedent beliefs.

Note that the Permissive Principle is very much like "inference to the best explanation" (IBE), but directed inwards. One of the implications of (IBE) is that when theoretical observation is silent or uncertain on the existence of certain entities, we should determine whether those entities exist in the way that best accords with our confidence in our antecedent theories. (IBE) is of course not beyond doubt, and neither is (PP). But I claim that the Permissive Principle is no worse off, from the point of view of epistemic responsibility, than inference to the best explanation.

The objector may grant that if we accept (PP) then we should accept that freedom is represented in experience.[107] But doesn't (PP) license incredible conclusions about the contents of experience? I think that it does not.

Consider the humors. There is in fact no such thing as the black bile traditionally associated with the melancholic temperament. Yet it was once widely believed that there was such a thing, and people took it to be part of the contents of their experience. Does (PP) entail that their view is correct? It may. Assuming that they are antecedently confident that they suffer from black bile, *and* there is no way of justifying the existence of black bile except if it is represented in experience, *then* they ought to be confident that the black bile figures in the contents of their experience. The objector claims that this is an absurd consequence of (PP).

I accept the consequence, but not its absurdity. If one is antecedently confident that there is black bile, and the other conditions of (PP) obtain, then one *ought* to conclude that black bile figures in the contents of ones experience. The problem is that the hypothesis that there is black bile is a dubious one. One should not accept this belief in the first place. But if one does, then the conclusion licensed by (PP) is a reasonable one to draw.

But perhaps this reasoning goes through only because inductive reasoning is not, after all, silent on the theory of the humors, as it is on the instantiation of the freedom relation. Consider then a divine being, coupled with the assumption that empirical evidence is silent on the existence or non-existence of such a being. Assuming that someone is antecedently confident that there is a divine being, *and* that there is no way of justifying the existence of such a being unless it is represented in experience (if, for example, one discounts all *a priori* arguments and the testimonial evidence), *then* one ought to be confident that a divine being figures in the contents of their experience. And the objector claims that *this*, at least, is an absurd consequence of (PP).

---

[107] As indeed we should, as I will explain below.

Again, I accept the consequence, but not its absurdity. It seems to me that much religious thought involves precisely this kind of reasoning, and it seems to me perfectly respectable from the point of view of rationality. Many find the *premises* of this line of reasoning objectionable, especially the first premise, but I do not see anything unreasonable in drawing the conclusion once one has accepted the conclusion, and so no objection here to (PP).[108]

If we accept (PP), then, conjoined with (JE), it yields an argument that freedom is represented in experience. Of course, the conclusion of this argument remains in a certain way hypothetical. For a given epistemic agent satisfies (PP) *only if* he is antecedently confident that the freedom relation is instantiated. If one believes this conditions to be met, *then* one ought to believe that freedom is represented in experience.[109]

I myself find this conditions to be met in my own case, and I therefore conclude that freedom is represented in my experience. In this respect, my method is somewhat similar to the old introspectionist method, whose conclusions ultimately depend on the agent's own beliefs. But my reliance on introspection is rather more indirect: I find that my antecedent confidence *entails* that my experiences have certain contents. I do not, like traditional introspectionists, propose to introspect those experiences directly.

What if someone is antecedently doubtful that the freedom relation is instantiated? Then this argument will not move him. But there is nothing surprising in this. It is merely a recurrence of a strain that has already appeared here. The task of the present essay is not to convince the skeptic about freedom that the freedom relation is instantiated. It is rather to show that we can answer the skeptic about freedom to our own satisfaction.[110]

---

[108] Notice however that (PP) claims something stronger, namely that it would be *unreasonable* to *not* draw the conclusion if one already accepted the premises. I am inclined to think that this stronger conclusion is correct. But it is tempting here to retreat to a *Doubly Permissive Principle* (DPP), which is like (PP) but for substituting an *epistemic permission* ('may') in where (PP) has an *epistemic recommendation* ('ought'). An interesting question, which I will not pursue here, is how the present theory would look if we were to accept (DPP) over (PP). I think that much of the formal structure would remain unchanged, but that the spirit of the view would become much different, closer in spirit to the view of James 1884.

[109] One might question the place where I have put the 'ought' in this hypothetical. For we might instead put the 'ought' before the conditional, so that (PP) reads: It ought to be the case that, if one satisfies (i) and (ii), then one is confident that R figures in the contents of one's experience. This would be to "wide-scope" (PP), to have it govern a "combination of attitudes" in the language of Broome 1999. It is not entirely clear, however, what the *motivation* for wide-scoping (PP) is. This stratagem is usually introduced as a method of blunting the allegedly implausible consequences of hypothetical principles like the instrumental principle, which seem to allow one to "bootstrap" one's way into reasons one does not antecedently have. But as I have argued in the foregoing, I am not at all sure that such bootstrapping – at least in the case of (PP) – is an implausible consequence. It does seem reasonable to grant, however, that if I am mistaken about this, then the wide-scoping stratagem *may* offer a way of eliminating these consequences of (PP) while retaining something of its spirit.

[110] But what if someone has an experience as of the freedom relation *not* obtaining? (It may be difficult to say "what it is like" to have such an experience, but perhaps no *more* difficult than the sometimes reported experience as of the *absence* of a divine being). It depends. If we grant heavy weight to introspective evidence, then this experience should defeat one's antecedent confidence in the belief that stands in the freedom relation to some actions. If we do not, then perhaps one should discount this experience as

This answer, however, is not yet complete. For the perceptual thesis to be true, then freedom must actually be *perceived* by agents, not merely represented. And the question of whether freedom is perceived cannot be decided merely by *a priori* methods alone, even when we accept the kind of epistemic permissiveness that (PP) licenses. How this question *is* to be decided is the topic of the next section.

4.9 Freedom is perceived

To say that freedom is represented in experience is not yet to say that it is perceived. For many representations fail to constitute genuine perceptions. When I hallucinate a dagger before me, I do not perceive a dagger, for there is no dagger there to be perceived. Indeed, even if there *happens to be* a dagger present, my hallucination does not constitute a perception, for the representation does not seem to *depend* upon the existence of the dagger in a suitable way.

Speaking more generally, the perception relation involves what we might call *external conditions*. As a first pass, we might take these to be as follows:

(ECP) S perceives that x is F *only if*
- (i) S represents x as being F
- (ii) x is F
- (iii) S's representation *depends on* x's being F

Freedom is perceived, then, only is the freedom relation *is* instantiated and our representation of freedom depends on its being instantiated. How do we decide whether this is so?

There is an obvious problem with deciding whether condition (ii) is met. For to decide this question, in the present case, is just to determine the fundamental question in the metaphysics of freedom: is the freedom relation ever instantiated? The problem is redoubled since, on my view, our only justification for believing (ii) arises from precisely the account that we are now trying to evaluate. So we seem to be stuck.[111]

There is however a sensible way forward. Let us hold in abeyance the question of whether the freedom relation is in fact ever instantiated. What we must go by then, is not whether the freedom relation is actually instantiated in some given case, but whether our *considered judgment* is that the freedom relation is instantiated in some given case. For we do have considered judgments about when someone is or is not free to, for example, raise his arm. It was precisely those judgments that allowed us to adjudicate the various

---

misleading, and take the true contents of one's experience to be given by one's antecedent confidence coupled with the implications of (PP). If one does have such an experience and adopts the first view of introspective evidence, then one will not find the present argument convincing. In this sense, while my argument does not *rely* on introspection, it may be *undermined* by it.

[111] A similar problem arises for condition (iii), since if one's representation depends on some thing's being F, then *a fortiori* that thing is F. Indeed, because of this, condition (ii) is strictly speaking redundant.

attempts to reduce the freedom relation in Chapter 2. What we should go on, in evaluating whether (ECP) is satisfied in the case of freedom, is our representations of freedom on the one hand, and our considered judgments about when the freedom relation is instantiated, on the other.

With those as our data, how do we determine whether (ECP) obtains? With the question of whether the freedom relation is in fact instantiated bracketed, our question reduces to whether our representations of freedom depend on the instantiation of the freedom relation, in cases where our considered judgment is that the freedom relation is instantiated.

What exactly does the dependence relation demanded by condition (iii) come to? Sometimes it is said that for a perception to occur, an observer's experience must be *caused by* the fact that the perceived state is instantiated.[112] This seems to cause trouble, however, when conjoined with the conjecture that freedom does not stand in causal relations at all. If that is true, then this would seem to rule out the possibility of perceiving freedom altogether.

I am inclined to think, however, that this reflects a problem with this formulation of dependence rather than with the perceptual view of freedom. For *many* putative objects of perception are such that they do not seem to stand in causal relations. Color, for example, if we grant that it is not reducible to its causal basis, is not something that stands in causal relations. Yet color is a paradigmatic object of perception. It *may* be that the causal requirement this tells against any non-reductive view of sensible qualities like color and, on my view, freedom. But this point is a contentious one, and we would do better for now with some more neutral formulation of dependence, one which does not embroil us in this particular difficulty.

I think a somewhat weaker while still robust version of the dependence requirement holds that the experience stands in a relation of *counterfactual dependence* to the fact represented. On this view:

(CD) S's experience of x being F is a perception that x is F *only if* if x were not F, then S would not have that experience

If this is correct, then the question of whether freedom is perceived turns on the following question: is the experience of freedom such that it normally satisfies (CD)?

This question cannot be decided wholly by a *priori* methods. It posits, after all, a relationship of counterfactual dependence between distinct existences. If this dependence is realized, then its realization is a contingent and *a posteriori* matter. Nonetheless, I think that it is plausible that this dependence is realized. Let me say, in several steps, why I think that this is so.

---

[112] Grice 1961.

First, note that, on my view, the experience of freedom is part of the normal content of bodily awareness. When one is aware of some part of oneself, such as a limb, one is typically at the same time aware of being free to move that limb. So there is a counterfactual dependence between the experience of freedom and freedom itself[113] *if and only if* there is a counterfactual dependence between normal bodily awareness and facts about freedom itself.

Second, observe that there is normally a *correlation* between normal bodily awareness and being free to act in various ways. A normal agent in normal circumstances enjoys full bodily awareness and is free to move in a variety of ways: to lift his arm, his leg, and so forth. In contrast, agents *without* freedom seem to also be deprived of normal bodily awareness. *Paralytics* are clear examples of agents who lack both the freedom to move their limbs *and* awareness of those limbs, and this seems to be true whether this paralysis is temporary (as when one's leg "falls asleep") or permanent.

Third, note that this correlation does not seem to be *accidental*. The absence of bodily awareness and the absence of freedom seem to be *explained by* the same underlying facts, namely deficiencies in the underlying physiology. This correlation is not necessary; there *could* be agents who lack bodily awareness without lacking freedom, and vice-versa. Indeed, as I will note presently, such cases are actual. Nonetheless, this correlation has *grounds*: its obtaining does not seem to be a coincidence.

This last point is the crucial one for our purposes. The fact that this correlation is a well-grounded one is what allows it to play its role in making (CD) true. Consider now your right arm. Normally, it is true both that you enjoy full bodily awareness with respect to that arm *and* that you are free to raise that arm. If the foregoing is correct, then this awareness constitutes a perception of freedom *only if* it is true that, if you were *not* free to raise your right arm, then you would *not* enjoy normal bodily awareness in it.

The question of whether this counterfactual is true turns on the extent of what we might call *hallucinations* of freedom: cases where one seems to be free though one, in fact, is not. There are innumerable mundane examples of such cases. For freedom, as we have said, can be undermined by extrinsic factors, and these factors will not normally themselves be objects of bodily awareness. Almost any obstacle can be sufficient to prevent one from being free without thereby destroying the sensation of freedom. So on the present view constrained agents are in a hallucinatory condition, one which is normally made manageable by their inductive knowledge that the deliverances of bodily awareness, in these cases, are not to be trusted.

But are hallucinations even more widespread than this? The extent of such hallucinations is, I think, not yet fully understood. Much of the recent empirical work on our consciousness of our bodies and its relationship to freedom is relevant here.[114] If, for

---

[113] Or rather between this experience and our *considered judgments* about when freedom obtains, in light of the considerations given above. I will henceforth leave this qualification implicit.

[114] Recall that what we are going by is not the obtaining of the freedom relation itself, but by our *considered judgments* about when the freedom relation obtains. This is one place where the "considered"

example, the decision to move one's arm normally precedes one's awareness of that decision, *and* one is no longer free to *not* move one's arm after that decision has been made,[115] then our ordinary experience is shot through with a kind of temporal hallucination which represents us as enjoying freedom for longer durations than we in fact do. This is why I say that the truth of (CD), and so the truth of (PT), remains uncertain and subject to disconfirmation.

To say this, however, is not to say that this view is *improbable*. Indeed, in light of the extant evidence, (CD) seems to me very probable. For while recent work has revealed unexpected hallucinations, it has not made anything like a case that such hallucinations are more prevalent than veridical perception, let alone that they obtain *globally*.[116] So I am inclined to conclude, in light of the present evidence, that (CD) is indeed true. This is why I claim that, in light of the other considerations adduced here, freedom *is* perceived.

4.10 The epistemic problem revisited

In Chapter 1 I asked the following question:

> Do we have any *evidence* that the freedom relation is instantiated?

We are now in a position to answer that question. In this chapter I have argued that (PT) is true, and hence that the freedom relation is something that agents normally perceive. So we ought to answer this question in the affirmative: agents *do* have evidence that the freedom relation is instantiated, namely the evidence provided by perception. And, if the arguments of Chapter 3 are correct, then this is the *only* evidence we have that the freedom relation is instantiated. So the conclusion of the previous two chapters is that evidence that the freedom relation is instantiated is provided by, and *only* by, perception.

To say that there is evidence that the freedom relation is instantiated is not to say what weight we ought to bestow upon this evidence, nor whether or when this evidence is defeated. There are a host of epistemological problems that remain even once we have accepted (PT). These problems can be divided into two kinds.

One kind of problem concerns perceptual justification quite generally. How is immediate perceptual justification even *possible*, and under what conditions does it obtain? These are important questions. I will not attempt to answer them here, but the form a final perceptual account of freedom will depend on how we resolve them. In this sense, a

---

qualification is relevant. These experiments, when convincing, *change* our judgments about when the freedom relation obtains. And it is these better informed judgments that we ought to go by in evaluating the correlation between our experience of freedom and the freedom relation itself.

[115] As is purportedly shown by Benjamin Libet's experiments. For a recent evaluation of these experiments, see Mele 2006, Chapter 2.

[116] The distinction between local and global hallucination is not frequently made in this scientific literature. It seems not to be made, for example, in Wegner 2002. My point here is simply that once we have made this distinction, the results of these experiments are, while still surprising, somewhat less *radical* than they are sometimes taken to be.

complete defense of the view stated here awaits an account of the justification that arises from experience more generally.

Another kind of problem is specific to the case of freedom. It is often claimed that justification for believing that the freedom relation is instantiated would also be justification for believing some other hypothesis, notably the hypothesis that *determinism* is false; it is sometimes claimed that this consequence, in turn, is an incredible one. This is just the problem of whether freedom is *compatible* with determinism, presented in its epistemic guise. Once we have accepted the perceptual thesis, we are I think in a better position to evaluate this problem. This is what I will try to do in the next chapter.

The Possibility of Freedom

Chapter Five: Compatibilism

5.1 What is compatibilism?

Thus far I have been postponing the contentious question of whether freedom is compatible with determinism, with the promise that we will be in a better position to address this question once we have a better grasp on the ontology and the epistemology of the freedom relation. Now it is time to make good on that promise.

The term *compatibilism*, though familiar, is not transparent. It is used in a variety of different senses by different authors. Sometimes these differences are made explicit; at other times, they remain implicit, and so engender confusion. Making clear exactly what compatibilism is, and how it is to be formulated, will be the business of the next three sections.

Let us begin at a very general level. To be a compatibilist with respect to some fact and some hypothesis is to hold, of that fact and that hypothesis, that they stand in *the compatibility relation* to each other. To make clear what exactly is meant by compatibilism in the present case it is necessary and sufficient to say what the two relata are, and to explain what it is for the compatibility relation to hold between them.

The fact that we are interested in here is the instantiation of *the freedom relation*. Very often the term "compatibilism" is used with some other fact as the first relatum, for example the instantiation of *moral responsibility*. This terminological ambiguity is harmless provided that we make explicit, when we say we are compatibilists, what the first relatum of the compatibility relation is. And that is what I have done here: the first relatum is the instantiation of the freedom relation, the very relation that has been at issue in the previous four chapters.

The hypothesis that we are interested in here is *determinism*. Doubts about the coherence of the debate over compatibilism often arise from the thought that determinism is somehow a confused doctrine: "a name for nothing clear," says Austin.[117] I have my own doubts about the coherence of some parts of this debate, which will emerge presently, but they do not turn on any doubts about the cogency of determinism.

I will take determinism to be a property of a world. A world has this property *just in case* a complete statement of the laws of that world conjoined with a complete description of the state of that world at any one time *entail* a complete description of the state of the world at every other time. For example, our world is deterministic just in case a complete statement of its laws conjoined with a complete description of its state at 7 am this morning entail a complete description of the state of the world at every other time, for example at 7 am tomorrow morning. There are a number of contentious concepts that figure in this definition: the concept of a world, and of a complete state of the world, and

[117] Austin 1956, p. 231.

of the laws of that world. A more thorough discussion of determinism would require making explicit how exactly each of these terms is to be understood.[118] But for present purposes this statement of determinism is I think both adequate and, *pace* Austin, perfectly clear.

It is often noted that determinism is a *scientific* hypothesis. I think this is in some intuitive sense correct, though it is difficult to make that sense explicit in a neutral way. One way of delimiting the class of scientific hypotheses is by saying that all of them are *contingent* and *a posteriori*. But, as will become clear, both the contingency and the a posterioricity of determinism are contentious questions. As a provisional way of underscoring this aspect of determinism, we may define "scientific" operationally. Determinism is a scientific hypothesis in the sense that such a hypothesis is one on which the deliverances of the sciences may have some bearing. Determinism is scientific in this sense because it is appears to be the *sort* of hypothesis that present or future physics might confirm or disconfirm.[119] This aspect of determinism, as will become clear, is central to the motivation of compatibilism.

So these are our relata: the freedom relation, on the one hand, and determinism, on the other. To be a compatibilist, in the sense we are interested in, is to hold that this fact and this hypothesis stand in the compatibility relation to each other. But what is the compatibility relation? This question is seldom asked, and it is surprisingly difficult to answer. In fact, there are two very different ways of answering this question. Distinguishing these varieties of compatibility, which I will do over the next two sections, will take us a long way towards understanding how exactly compatibilism is to be understood, and how it is to be defended.

5.2 Metaphysical compatibilism

To say that the freedom relation is compatible with determinism is to say that their combination is, in some sense, *possible*. But what sense of possibility do we have in mind? One natural answer is to say that it is *metaphysical possibility* that we have in mind, and to take the compatibilist to be someone who holds that the freedom relation and determinism are compossible in precisely this sense: that the conjunction of the freedom relation and determinism is metaphysically possible. In this section I will explore some difficulties for this conception of compatibilism.

Let us state metaphysical compatibilism as follows:

(MCOMP) It is metaphysically possible that the freedom relation is instantiated and that determinism is true

---

[118] Precisely this (and more) is done in Earman 1986. One crucial point, emphasized at length by Earman (pp. 4-10), is that determinism is a thesis in the first place about a world, rather than about an *observer of* a world. So the advocate of determinism is not straightaway committed to any thesis about an observer, say about what an omniscient observer (a "Laplacian demon") could *know* about the future in such a world.

[119] For more on the relationship between physical theories and determinism, see again Earman 1986, especially Chapters 2-3 and 10-11.

Is (MCOMP) a good conception of compatibilism; that is, one that is endorsed by anyone whom we intuitively would wish to class as a compatibilist?

One initial problem is how we are to classify the following sort of person. Consider someone who holds that it is impossible that the freedom relation is ever instantiated. Perhaps he thinks the very idea of such a relation is incoherent, or perhaps he thinks that, while such a relation is coherent, it is necessarily never realized. Such an agent comes out as *denying* compatibilism in the present sense.

This seems to be correct. Let us reserve the term *impossibilist* for anyone who asserts that the freedom relation necessarily is never instantiated.[120] The question of compatibilism and its denial does not arise for such a person, for there is necessarily not any relation *to be* compatible or not compatible with determinism. To capture the debate between compatibilism and its denial, we should restrict our attention to those who accept that it is possible that the freedom relation is instantiated; that is, to the *possibilists* about the freedom relation. The *compatibilists*, then, are a proper subset of the *possibilists*.

But there is a more serious problem for (MCOMP). Imagine that someone holds the following view. He asserts that the freedom relation is possible (indeed, even that it is actual), *and* that the question of freedom is wholly independent of the question of determinism, so that in some intuitive sense the truth or falsehood of determinism has *nothing to do* with whether the freedom relation is instantiated. Such a person, intuitively, ought to count as a compatibilist. But imagine that this person also holds that *determinism* is impossible. Perhaps he holds, on broadly empirical grounds, that the laws of our world are indeterminstic, *and*, on broadly metaphysical grounds, that the laws of nature are necessarily true (that "physical necessity is necessity *tout court*"[121]). Such an agent *denies* (MCOMP), in virtue of denying its second conjunct. So he does not, on the present conception of compatibilism, come out as a compatibilist. There seems to be a problem here.

Let me be very clear about what that problem is. I do not think that the denial of determinism's possible truth is a particularly widely-held or even a particularly plausible theory. But it does seem a theory that is coherent, not entirely absurd, and one that *could* be held by someone whose views on *freedom* were in line with those who are considered compatibilists. The problem is that (MCOMP) seems to saddle one who wishes to be a compatibilist with auxiliary metaphysical commitments that intuitively have nothing to do with compatibilism. (MCOMP) seems not to express compatibilism itself, but rather a particular conjunction of metaphysical views that happen to inhabit the neighborhood of compatibilism.

What is the upshot of all of this? I do not claim that there is no coherent doctrine of metaphysical compatibilism. That would be premature, for there are a number of ways of

[120] This useful term is due to Vihvelin 2007.
[121] Kripke 1980, p. 164.

formulating this view that I have not even considered here.[122] But I think our difficulties here make it reasonable for us to pursue a rather different conception of compatibilism. As I will argue in the next section, this conception avoids the difficulties we have encountered here, as well as yielding an explanation of how metaphysical views of compatibilism manage to *approximate* to the correct conception of compatibilism.

5.3 Epistemic compatibilism

I think our problem in even *formulating* metaphysical compatibilism arose from the following fact: compatibilism is not best understood a thesis about the *metaphysics* of freedom and determinism at all. It is rather best understood as a thesis about their *epistemology*.

It is important to be clear about what these claims about how compatibilism is "best understood" come to. "Compatibilism" is a technical term that can be and is used in a variety of ways. Many authors are more or less explicit about what they mean by it, often defining it by the possibility thesis that I have called metaphysical compatibilism. Yet these definitions are in the service of a *conception* of the relationship between freedom and determinism, one which is sometimes elusive but of which its proponents have an intuitive grasp. In saying that compatibilism is not "best understood" in terms of metaphysical compatibilism, I am saying that this conception is ill-served by this definition. There is another way of making compatibilism explicit, one which better captures the conception that guides its advocates. And that is what I will try to spell out in this section.

Let us first try to state *epistemic compatibilism* in a way that follows our statement of metaphysical compatibilism:

(ECOMP) It is epistemically possible that the freedom relation is instantiated and that determinism is true

Is (ECOMP) endorsed by all and only those whom we would intuitively count as compatibilists? Let us begin with the *epistemic impossibilist*. This is someone who holds

---

[122] One idea is that the compatibilist holds that it is the *essence* of the freedom relation to obtain independently of the truth of determinism, in something like the sense of essence in Fine 1994. For the present case does seem to be one, like Fine's, in which talk of necessity and possibility seems inapt to capture the salient issues. But there are problems with this strategy. For one thing, the essence of a thing is normally an *intrinsic* matter, which the question of whether something is in a deterministic world or not seems to be an *extrinsic* matter. For another, it is doubtful that the freedom relation has an essence that we might make explicit (that is, of which we might give a *real definition*) if I am correct that, as I argued in Chapter 2, this relation is a primitive one.

Another idea, similar in spirit, appeals to the "in virtue of" relation. Have the metaphysical compatibilist assert that *if* determinism were false (which may be, in light of the above, a counterpossible), then this is not so *in virtue of* the fact that agents are free to perform actions they do not perform. One problem for this proposal is that it seems to overgenerate. A typical incompatibilist may accept this conditional: he may for example claim that the explanation for determinism's falsehood lies in the nature of the laws of our world, and that human freedom is merely an *upshot* of this physical fact, and so a "spandrel" from the point of view of the universe.

that he *knows* that the freedom relation can never be instantiated, perhaps because the very concept of such a relation is incoherent (like a round square). Such a person *denies* (ECOMP), for he denies that the instantiation of freedom *is* a genuine epistemic possibility. And again it seems correct to exclude this sort of view from the class of the compatibilist views, for the debate over compatibility is one that arises only for one who is willing to accept the possibility – in this case, the epistemic possibility – that the freedom relation is sometimes instantiated.

But what of a figure whom we should intuitively wish to count as a compatibilist but who also holds that he *knows* that determinism could not be true? If there were such a person, then (ECOMP) would face precisely the same problem that (MCOMP) faced on this score. But I think that there is a genuine asymmetry here. Someone who holds that physical necessity is necessity *tout court* might reasonably count as a compatibilist, for the question of the modal status of the laws seems wholly orthogonal to the question of compatibilism. But the same is not true in the present case. For it is precisely the idea that we cannot know the truth of scientific hypothesis "from the armchair" that motivates the compatibilist. So while there may be those who hold that we know that determinism cannot be true, these are not figures whom we should wish to count as compatibilists. For it is precisely the point of the compatibilist view that the truth of scientific hypotheses is an empirical question, one that, from the armchair, remains open.[123]

But there is a rather different problem that arises for (ECOMP). Consider a figure who endorses all the standard arguments for incompatibilism and holds that he knows, on these grounds, that determinism is false. But imagine that he also harbors a reasonable amount of doubt about the limits of his philosophical insight, as well as recognizing that experiments confirming the truth of determinism might, for all he knows, be announced in tomorrow's scientific journals. He holds that, *if* such an announcement were to be made, he would become a compatibilist. So he does not deny that it is epistemically possible[124] that the freedom relation is instantiated and determinism is true. So he endorses (ECOMP). Yet he is not *now* ("nor has he ever been") a compatibilist, though he would under certain circumstances become one. But this account of compatibilism wrongly implies that he is now a compatibilist. So we must reject (ECOMP) on the grounds that it is so weak as to be endorsed by the opponents of compatibilism.[125]

---

[123] We should be careful, however, not to exclude another sort of figure. This is someone who ought to count as a compatibilist but who holds that determinism is actually false, and that its falsehood is confirmed to a sufficiently high degree for us to *know* that determinism is false. Indeed, it seems not improbable that this is our actual epistemic situation. In order to not exclude this sort of figure, we must construe epistemic possibility as *undemandingly* as possible. Informally, we can say that some proposition p is epistemically possible for some agent S just in case he cannot *rule out* the truth of p. A more careful statement of this idea would require a more thorough discussion of epistemic possibility than I will undertake here, in particular of the *ruling out* relation. Nonetheless, I think we have a sufficient grasp of this sort of epistemic possibility to guide our thinking in what follows. Indeed, it is what will presently lead me to *reject* (ECOMP) on the grounds that it overgenerates.

[124] In the undemanding sense required to avoid the problem raised the prior footnote.

[125] The view described in this paragraph is precisely that of Peter van Inwagen; see van Inwagen 1983, p. 223.

What has gone wrong? I think that we were correct to turn to an epistemic formulation of compatibilism, but incorrect to state compatibilism in terms of epistemic *possibility*. Let me now state an epistemic formulation of compatibilism that does capture all and only those people whom we should count as compatibilists. This formulation will make no explicit mention of possibility, be that possibility metaphysical or epistemic. Finally, I will explain why formulations of compatibilism in terms of metaphysical possibility manage to *approximate* to the correct formulation of compatibilism.

I think the proper formulation of compatibilism is to be stated in terms of *evidence*. On this formulation, compatibilism will turn out to be a negative thesis. It does not assert a certain thesis about possibility. It rather *denies* a certain thesis about evidence:

(ECOMP*) It is *not* the case that, if some body of evidence E is evidence that the freedom relation is instantiated, then E is evidence that determinism is false

Since this statement is a negative one, it will also be useful to have a label for *epistemic incompatibilism*, the position whose *denial* is central to compatibilism:

(EINCOMP) If some body of evidence E is evidence that the freedom relation is instantiated, then E is evidence that determinism is false[126]

This statement of compatibilism and its denial introduces some new concepts: the concept of a *body of evidence*, and an *evidential relation* that holds between bodies of evidence and propositions. As at other points in this essay, I will try to remain ecumenical on a variety of questions about the nature of evidence, and work instead with an intuitive notion of bodies of evidence and the evidence relation.[127]

How does (ECOMP*) fare with respect to the various intellectual characters we have been considering? Again, we must consider a certain sort of imposssibilist, the one who denies that there *could* be evidence that the freedom relation is instantiated, perhaps because the freedom relation is incoherent and there could be no evidence that that an

[126] What if an omniscient and truth-telling divine oracle were to announce: "Determinism is true *and* the freedom relation is instantiated." Wouldn't *that* be a counterexample to (EINCOMP), and shouldn't anyone admit that possibility, and so isn't (EINCOMP) clearly false? We must be careful here. Such a pronouncement would indeed be evidence that the freedom relation obtains and would not be evidence that determinism is false. But not everyone should admit this possibility. The incompatibilist *denies* that such a pronouncement is a genuine possibility (it is an *epistemic possibility* for him, but that is not what is at issue; what is at issue is whether it is *metaphysically possible* that there be such evidence. I discuss this point further in footnote 128). The incompatibilist simply *denies* that a genuinely omniscient and truth-telling oracle could make such a pronouncement, and this should not be surprising, for such an oracle would be announcing, in effect, that incompatibilism is false.

[127] See Kelly 2006 for a summary of historical and recent work on evidence. I do not mean to downplay the significance of giving a more informative account of evidence. Indeed, if my formulation of compatibilism is correct, then such an account may have a heretofore unexpected bearing on questions of freedom and determinism.

incoherent relation is instantiated.[128] Such an impossibilist will count (ECOMP*) as trivially false (since he counts (EINCOMP) as trivially true). So we do not need to supplement (ECOMP*) to capture the idea that the question of compatibilism arises only for someone who is willing to grant that there could be some evidence that the freedom relation is instantiated.

How about the compatibilist whom (MCOMP) counted as an incompatibilist, and the incompatibilist whom (ECOMP) counted as a compatibilist? (ECOMP*) classifies each of these correctly. The compatibilist who holds that determinism is necessarily false will *accept* (ECOMP*), since he thinks that questions about freedom and determinism are unrelated to each other; his views about the necessity of indeterminism do not even come into the picture, as they should not. On the other hand, the incompatibilist who holds that determinism might be confirmed and that his views might change will *deny* (ECOMP*) (that is, he will accept (EINCOMP)) given his epistemic situation *now*; to say that he might change his views is just to say that he might come to reject (EINCOMP).

There is another sort of problem however that arises for (ECOMP*), however. Imagine that someone intuitively counts as a compatibilist, rejecting all arguments for incompatibilism as unsound. But let us also imagine that he reasonably believes (perhaps a reliable divinity told him) that evidence that the freedom relation is instantiated will be revealed to him just in case the world is indeterministic. This is not because of any connection between freedom and indeterminism, but simply because his evidence is being "managed" in a certain way (perhaps by the divinity himself). Such a person will *reject* (ECOMP*), because he holds these bodies of evidence to be correlated.

This problem is I think to be avoided as follows. The correct statement of epistemic compatibilism is as follows:

(ECOMP**) It is not the case that any evidence E that the freedom relation is instantiated is evidence that determinism is false *just in virtue of* being evidence that the freedom relation is instantiated

The character we have just considered rejects (ECOMP*), but he accepts (ECOMP**). His evidence that the freedom relation is instantiated is not evidence that determinism is false just in virtue of being evidence that the freedom relation is instantiated. It is rather evidence that determinism is false partly in virtue of what the divinity is told him. So I think we may avoid this problem by reformulating (ECOMP*) as suggested in (ECOMP**).[129]

With that alteration in place, we have I think arrived at an adequate statement of epistemic compatibilism. The epistemic incompatibilist asserts that there is an *essential*

---

[128] This 'could' expresses metaphysical, rather than epistemic, possibility. So the assertion of compatibilism implicitly asserts a metaphysical thesis after all, but this assertion does not concern the freedom relation or determinism, but rather *evidence* for the freedom relation and determinism. Since the *object* of the compatibilist's claim is in the first place evidence, "epistemic compatibilism" remains an apt name for it.

[129] This point applies also to our statement of epistemic incompatibilism, *mutatis mutandis*.

connection between evidence that the freedom relation is instantiated and evidence that determinism is false, because of which any evidence for the instantiation of the freedom relation is, *ipso facto*, evidence for the falsehood of determinism. The epistemic compatibilist *denies* that there is any such connection.[130]

We can also now say why metaphysical compatibilism seemed to be approximately true. This is because (MCOMP) *entails* (ECOMP**). If it is metaphysically possible that the freedom relation is instantiated *and* determinism is true, *then* there is no essential connection between evidence for freedom and evidence for the falsehood of determinism, so (ECOMP**) true. So anyone who is a metaphysical compatibilist will *also* be an epistemic compatibilist. But this is not because metaphysical compatibilism is an apt statement of compatibilism; as we have seen, it is not. It is rather because it expresses a metaphysical thesis with an interesting property: *if* someone accepts that thesis, *then* he will be a compatibilist. But the converse is not true, for accepting (ECOMP**) is itself *sufficient* for being a compatibilist.

Now that we have said what compatibilism *is*, we are finally in a position to argue about whether it is *true*.

5.4 The consequence argument

Debates over compatibilism and incompatibilism often begin with claims that one of these positions is more *intuitive* than the other. It is difficult to know how to assess these claims, and I will not make any such claims here.[131] Nonetheless, there is a feature of these debates that an adequate account of them might be expected to explain. Whether incompatibilism is more or less intuitive than its denial, the debate over the question of compatibility typically involves the incompatibilist giving *arguments* for his position, and the compatibilist trying to show why those arguments are unsound. Debates over the question of compatibility, then, involve what we might call a *dialectical asymmetry*.

The explanation of this asymmetry is not, I think, merely sociological. It rather arises from a feature of the incompatibilist's claim as we have described it. The incompatibilist asserts that there is an essential connection between the evidence for two claims that are at least *nominally* distinct, a connection such that any evidence for one is evidence for the other. The compatibilist denies any such connection. We might compare their positions to the *color-incompatibilist* and the *color-compatibilist*. The color incompatibilist asserts that:

---

[130] What of the following character? He believes in some intuitive sense that freedom and determinism are compatible, but he also believes that a divinity would never allow them to be coinstantiated in a world – not because of anything in their natures, but simply because he happens not to like the idea of their copresence. This character assents to (ECOMP**): he thinks that evidence that the freedom relation is instantiated is evidence that determinism is false, but *not* just in virtue of the fact that it is evidence that the freedom relation is instantiated. So he comes out, on the present definition, as a compatibilist. I think that this is the correct result. This seems to me a case of a genuine compatibilist who happens to think, because of his esoteric beliefs about divine tastes, that freedom and determinism could never be coinstantiated.

[131] For one recent attempt to at such an assessment, see Nahmias et al. 2006.

(CINCOMP) Any evidence E that some object x is red all over is evidence that x is *not* green all over *just in virtue of* being evidence that x is red all over

The color-compatibilist, in turn, denies (CINCOMP). Were the color-incompatibilist called upon to defend his position, he could offer an argument demonstrating an *a priori* connection between the two bodies of evidence, as follows:

(1) E is evidence that x is red all over
(2) It is a priori that nothing can be red all over and green all over[132]
(C) E is evidence that x is not green all over

We may usefully think of the incompatibilist as trying to defend a similar sort of claim by establishing a similar sort of *a priori* connection. This is what explains the dialectical asymmetry between the incompatibilist and his opponent.

The color-incompatibilist's argument implicitly appeals to a principle which it will be useful to keep in mind in what follows. This is what we might call the principle of *a priori transmission*:

(APT) If E is evidence that p, and it is a priori that if p then q, then E is evidence that q *just in virtue of* being evidence that p[133]

The grounds for accepting (APT) are something we will look at more carefully below. First, let us look in detail at the incompatibilist's argument for (EINCOMP), being sure to make explicit its appeal to (APT).

The most compelling such argument for (EINCOMP) is what is commonly known as "the consequence argument." This is not really the name for an argument, but rather for a *kind* of argument, of which there are several developments.[134] The argument that I will consider is a rather standard development of this argument, but for my making explicit the epistemic operators that are needed for the argument to go through, if my formulation of incompatibilism is the correct one.

Here is the argument. Say that we have evidence E that some agent S was free to perform some action he did not perform. Let 'P' denote the proposition that S did *not* do

---

[132] As this example makes clear, the evidential connection need not be in any substantive sense *analytic*.

[133] A similar principle applies to cases where the intermediate premises are themselves supported by independent evidence, in which case we might say that the evidence for the first premise and the evidence for the intermediate premise or premises *jointly transmit* to the conclusion. The phenomenon of evidence transmission is in this sense a quite general one. (APT) represents what we might think of the purest and simplest case of this more general phenomenon, where there is no independent evidence but for that which is mentioned in the first premise. Since this is the only such principle that the incompatibilist's arguments need to appeal to, it is the only principle that I will discuss here.

[134] The name is due to van Inwagen 1983, who himself offers three different developments of this argument (see pp. 55-105). The formulation of this argument which I will soon present is essentially that of what van Inwagen calls his "first argument," but for my inversion of the direction of the argument and addition of epistemic operators.

A, 'L' the laws of the world, and 'H' the complete history of the world up to the time at which S did not raise his hand. Then:

(1) E is evidence that S was free to A
(2) It is a priori that, if S was free to A, then S was free to render P false
(3) It is a priori that, if S was free to render P false, and H and L entail P, then S was free to render H or L false
(4) It is a priori that, if determinism is true, then H and L entail P
(5) It is a priori that, if determinism is true and S was free to A, then S was free to render H or L false (by 2,3,4)
(6) It is a priori that, if S was free to render H or L false, then S was free to render L false
(7) It is a priori that S was *not* free to render L false
(8) It is a priori that either determinism is not true or S was not free to A (by 4,5,6)
(9) It is a priori that, if S was free to A, then determinism is false (by 8)
(C) E is evidence that determinism is false just in virtue of being evidence that S was free to A (by 1, 9, and (APT))

I have put this argument in terms of a specific case, but this has involved no loss of generality, since E, S, and A may represent *any* evidence, *any* agent, and *any* act. Therefore, if this argument is sound, then (EINCOMP) is true. So I am in agreement with the standard view that the truth of incompatibilism turns on the soundness of the consequence argument, so long as that argument includes the epistemic operators I have included, as well as the epistemic principle (APT).

It must be admitted, I think, that this argument is a compelling one. How is the compatibilist to respond to it?

5.5 The Lewisian response

We have already considered one way of responding to this and similar arguments.[135] This is by proposing a reduction of the freedom relation, for example to say that S is free to A just in case S would A if he chose to. If this reduction were true, then this argument would not be sound. For premise (7) seems to be false. It may be true of an agent that, *if* he were to choose to do something that would render the laws false, *then* he would break the laws. There is nothing surprising here, any more than it is surprising that *if* some object were to have a force greater than its mass by its acceleration, then Newton's second law would be false. So long as the laws ensure also that the *antecedent* of these conditionals will never hold, there is nothing here that conflicts with the inviolable governance of the laws.

But, as I have already said, there are two fatal problems with this response to the compatibilist argument. First, there *is* no reduction of the freedom relation – that is, no relation that is substitutable *salva veritate* for all occurrences of the freedom relation

[135] In Chapter 2.2.

(except for the freedom relation itself and its trivial transformations).[136] Second, even if there *seemed* to be such a relation, it is not clear how introducing it here would advance the compatibilist's case. For insofar as premise (7) of this argument is compelling, then it itself constitutes a counterexample to any substitute relation that would make this premise false. So there is no reduction of the freedom relation, and even if there seemed to be, any such reduction would be in this case unfit to play what I have called the *dialectical role* of a reduction.

So what we might fairly call the *traditional* response to the consequence argument will not work. A better response, however, was proposed by David Lewis.[137] Lewis's response to the consequence argument does not turn on any reduction of the freedom relation. Rather, it diagnoses an *equivocation* in the consequence argument. As will become clear, I think Lewis's presentation of his response is in some important ways *incomplete*. Nonetheless, it constitutes the beginning of a proper response to the consequence argument. Let me begin by introducing the Lewisian response roughly as Lewis himself presented it.

Lewis's response turns on a distinction between the "weak ability" and the "strong ability" to render a proposition false. I will speak instead of being *free to weakly* render a proposition false and being *free to strongly* render a proposition false. I use this terminology for two reasons. First, in order to stay consistent with the terminology introduced in Chapter 1 to mark the distinction between freedom and ability (or between "specific ability" and "general ability," respectively). This does no injustice to Lewis's point, for the context of the essay makes clear that it is the freedom relation which Lewis has in mind.[138]

The second reason is slightly more substantial. Lewis's terms carry the misleading implication that he is diagnosing an ambiguity in the freedom relation itself. This may seem to be at odds with my insistence on the indivisibility of the freedom relation. But this is merely a misleading implication of Lewis's terms. Lewis is not diagnosing an ambiguity in freedom, but rather in the technical notion of *rendering a proposition false*. Lewis distinguishes two kinds of *actions* that might be taken to deserve that name.

Lewis's distinction is a quite general one, but for present purposes we may restrict our attention to cases where the proposition rendered false is a *law*. As will become clear, that is sufficient to diagnosis where Lewis thinks the consequence argument goes wrong.

Consider first *strongly rendering a law false*. An action strongly renders a law false just in case that action either *is* or *causes* a law-breaking event. For example, if I move my hand faster than the speed of light, then my action *is* a law-breaking event. Or, if I throw a stone very hard, and so that stone thereby accelerates to a velocity greater than the

[136] This was the argument of Chapter 2.
[137] In Lewis 1981.
[138] Otherwise my presentation of the argument in what follows hews to Lewis's essay rather closely, so argumentative trappings (such as examples) should all be assumed to be Lewis's rather than mine.

speed of light, then my action *causes* a law-breaking event. In either case, I have *strongly* rendered the law that nothing can travel faster than the speed of light false.

Distinguish this from *weakly rendering a law false*. An action weakly renders a law false just in case (i) it does not *strongly* render a law false and (ii) if that action were performed, then a law would be broken. What such actions are there? In a deterministic world, almost any commonplace action that one is free to perform but does not perform is one that would weakly render a law false, if one were to perform it. Imagine our world is deterministic and that I am free to raise my hand but do not raise it. If I had raised my hand, then either a law would have had to been broken or the past would have had to been different. But, in the case where a law has been broken, it is not true that my action *is* or *causes* a law to be broken. Rather, the law-breaking event would be a *condition of possibility* of my raising my hand in the first place.

With that distinction in place, we can diagnose where the argument above goes wrong. On either sense of rendering a proposition false, premises (1) through (5) are all true. But at this point the distinction which Lewis makes comes into play. Consider first the case where it is *weakly rendering a law false* that we have in mind. In this case, Lewis says, we may grant that premise (6) – that it is a priori that if someone were free to weakly render the conjunction of history and a law false, then he would be free to weakly render a law false – is true. But premise (7), that it is a priori that no one is free to weakly render a law false, is not true. In a deterministic world where the freedom relation is instantiated, agents are regularly free to weakly render laws false.

Consider next the case where it is *strongly rendering a law false* that we have in mind. In this case, premise (7) is true. It is indeed a priori that no one is free to strongly render a law false. But in this case, premise (6) is false. It is not a priori that if someone were free to strongly render the conjunction of history and the laws false, then he would be free to strongly render a law false. For since it is impossible for anyone to strongly render a law false, if it were true that someone were free to strongly render the conjunction of history and the laws false, then this could only be because he was (somehow or other) free to strongly render the facts of history false.

Thus Lewis suggests that, once we are clear on the two ways of rendering a proposition false, the consequence argument fails. In order to avoid equivocation, we must specify *which* sense of rendering a law false we have in mind. But on *neither* reading is every premise of the argument true. The argument seems sound, then, only because it illicitly equivocates on the crucial notion of rendering a proposition false.

5.6 Two versions of the Lewisian response

Peter van Inwagen alleges that Lewis's response succeeds only because of superficial and dispensable aspects of this presentation of the argument.[139] If the proponent of the argument gets more *precise* about "rendering false," then he may reconstruct a version of the consequence argument against which the Lewisian response does not succeed.

[139] In van Inwagen 2004.

According to van Inwagen, we may carry out this reconstruction of the argument as follows. Everywhere delete the phrase "is free to render P false" and replace it with: "[is free to] arrange things in a certain way, such that his doing so, together with the whole truth about the past, strictly imply the falsity of [P]"[140] Once we make *this* substitution, alleges van Inwagen, the consequence argument remains sound and also invulnerable to the charge of equivocation.

The first question to be asked is whether this substitution illicitly imports precisely the point to be shown, namely the alleged truth of incompatibilism, into the premises.[141] To decide whether this is so, we need to decide whether a certain *freedom translation schema* between our ordinary talk about the freedom relation and van Inwagen's substitution is true. This schema is:

(FS) S is free to A just in case S is free to arrange things in a way such that the whole truth about the past is no different *and* S does A

For the revised consequence argument to work, (FS) must not only be true and *a priori*, but also must be clearly true even to those who are antecedently *agnostic* on the question of incompatibilism. Let us say (FS) is *transparently true* just in case it meets this standard.[142]

Transparent truth is a very high standard for any proposition to meet, and I do not know how to decide whether or not (FS) meets it. But I think we can here develop *two* versions of the Lewisian response, one for the case where (FS) is transparently true, the other for the case where it is not.

Let us say that (FS) is *not* transparently true. Then we need an interpretation of our way of speaking about freedom that explains why it seems to imply the sort of freedom to break the laws that it does, and why the consequence argument seems to be sound. And the Lewisian response is well-suited to play this role. These seemings are to be explained by an ambiguity that enters into our thinking about freedom when it is put in terms of rendering propositions false, an ambiguity between the freedom to *weakly* render a proposition false and the freedom to *strongly* do so. So understood, Lewis's response constitutes what we might call a *hermeneutic* proposal about our thinking about freedom.[143]

Let us say that (FS) *is* transparently true. Then the consequence argument *is* sound, at least if we also assume the truth of (APT), and the compatibilist is under a rather different

---

[140] Ibid., p. 346.
[141] This is the charge Lewis 1981 makes; see p. 296, footnote 5.
[142] Transparent truth is perhaps something very much like *analyticity*, though it would take some care to put the present point in those terms. For one thing, analyticity is in the first place a property of *sentences*, while transparent truth is a property of *propositions*. This is not to say that I have anything like an account of transparent truth: it should be rather taken as a placeholder for an account of propositions that are, in some intuitive sense, incontestably obvious.
[143] This term (and its antonym, "revolutionary," which I will introduce presently) is due to Burgess 1983.

burden. He needs to explain how we can continue to speak of freedom without committing ourselves to the consequence that our claims about freedom constitute evidence against the truth of determinism. In this case, the Lewisian response plays a rather different role. Instead of speaking of the freedom relation, we may speak instead of the *innocent freedom relation*, which is *exactly* like the freedom relation but for the fact that (FS) does not express a true claim about it. The innocent freedom relation is a way of speaking about freedom for those who do not wish to commit themselves to the claim that they possess evidence of the falsehood of determinism. It involves precisely the same freedom with respect to the laws as Lewis proposes that the *actual* freedom relation does. On this development, the Lewisian response is to be taken as a *revolutionary* proposal, a proposal to *change* our way of speaking about freedom.

This response is similar to a move that is not uncommon in the contemporary compatibilist literature. According to this move, what the compatibilist ought to do is to replace talk of freedom with talk of something that is "near enough to the folk's [conception of the freedom relation] to be regarded as a natural extension of it."[144] There is a crucial difference, however, between the present proposal and this sort of ersatzism. The ersatzist proposes a relation which does not have the freedom relation itself as one of its constituents, usually one of the failed analyses surveyed in Chapter 1. The proposal here is rather different: the proposal is to hew in our definition to the freedom relation itself but to then *subtract* from it the commitment to (FS).

This difference, however, also points to what may be a fatal difficulty for this revolutionary proposal. For it is not clear that this subtraction yields a determinate solution.[145] The incompatibilist may respond that there is *no such thing* as the freedom relation without a commitment to (FS), just as there is no such thing as the marriage relation without a commitment to spouses. Whether or not this point is ultimately sound, it indicates that the development of this sort of revolutionary proposal is a far from trivial task. It is also not a task that I will undertake here. I will, however, soon suggest a very different and I think preferable response that the compatibilist may make, if (FS) does indeed express a transparent truth.

Let us say, however, that we were to overcome these difficulties for the revolutionary proposal. What would we have accomplished? We would not have really given a *response* to the consequence argument at all. We would rather have adopted a way of speaking that *avoids* the implications of the consequence argument. The question then arises of why anyone would want to do this. If the consequence argument is sound, then why should we not just *accept* its implications? To answer *this* question, we need to ask what motivates compatibilism in the first place.

---

[144] Jackson 1996, p. 45. A longer citation of this passage occurs above in footnote 1 to Chapter 1.

[145] These sorts of worries about the vices of subtraction, as well as ideas about its possible virtues, were raised by Stephen Yablo in lectures at Princeton in March 2008.

5.7 The compatibilist principle

Earlier[146] I noted that part of what motivates compatibilism is the thought that determinism is in some sense or other a "scientific" hypothesis, but that it is difficult to say in precisely *what* sense this is so. Here I will try to do better. I have already argued that, on its best formulation, compatibilism is a thesis (or rather the denial of a thesis) about the relationship between certain bodies of evidence. I think that the principle that *motivates* compatibilism is also a certain principle about evidence.

At first pass, we might take the compatibilist to hold that:

(CP) Only *scientific* evidence is evidence for the truth or falsity of determinism

The burden is on the compatibilist to say what sense the qualifier "scientific" has here, and why the evidence typically adduced in favor of the instantiation of the freedom relation does not count as "scientific" in this sense.

Let us first try the working operational definition that we invoked earlier when this sort of question arose. We might say evidence is scientific just in case it is evidence that is given epistemic weight by contemporary scientists. Once we push this account, however, it is clear that it will not hold up. For we are faced with a familiar dilemma. On the one hand, we could go by what scientists *actually* give epistemic weight to. But in that case, we rule out by fiat the very possibility that contemporary scientists are failing to take into account some properly scientific evidence. On the other hand, we could go by what scientists *ought* to give epistemic weight to. But in that case, the compatibilist principle becomes trivial. For even the incompatibilist can agree that only the sort of evidence that scientists *ought* to give epistemic weight is evidence for the truth or falsity of determinism. He simply adds that scientists ought to give epistemic weight to our evidence for the instantiation of the freedom relation.

Another way to define "scientific" here is by adverting to the distinction between the *a priori* and the *a posteriori*. On this view, evidence is scientific just in case it is a posteriori evidence. But this definition both lets in too little and too much. Too little because much of what is naturally counted as scientific evidence is a priori: in particular, *thought experiments* often play an essential role in the evaluation of scientific theories. Too much because on some accounts, such as the one I defended in Chapter 4, our evidence for the instantiation of the freedom relation is itself a posteriori, and therefore "scientific" according to the present definition. So, in addition to its impoverished conception of scientific method, this definition may fail to exclude precisely the sorts of evidence for the truth or falsity of determinism which the compatibilist wants to exclude.

Here I want to sketch a better way of accounting for the role of "scientific" as it figures in the compatibilist principle. This account will be incomplete in various respects, and much more could be said about it. But it is, I think, sufficiently clear, compelling, and non-trivial for it to capture the motivation for compatibilism.

---

[146] In Section 1.

This account turns essentially on a distinction between the *first-personal* and the *third-personal* perspectives.[147] Informally, the first-personal perspective is a perspective on the world from a particular standpoint, with a determinate temporal and spatial location coupled with the phenomenology of "what it is like" to occupy that perspective. The third-personal perspective is a perspective on the world that *abstracts* away from any such determinate location and phenomenology; it is, in Thomas Nagel's phrase, the "view from nowhere."[148]

In terms of this distinction between *perspectives* we may make a distinction among two kinds of *evidence*. Let us say that evidence is *first-personal* just in case the possession of it requires a first-personal perspective. In short, someone who occupied no particular perspective would necessarily *lack* this sort of evidence. This class of evidence properly includes evidence that is essentially indexical, such as evidence that *I* am in the library, but such evidence does not seem to exhaust it. For example, our normal evidence that there is *pain* is paradigmatically first-personal: it would not be available to someone who did not occupy some particular perspective or other.[149] And let us say that evidence is *third-personal* just in case it is not first-personal.

We may then state the compatibilist principle as follows:

(CP*) Only *third-personal* evidence is evidence for the truth or falsity of determinism[150]

---

[147] I take these perspectives to be two ways of viewing the same world, and hence to stand in potential conflict, rather than incommensurable "points of view" which can never conflict. Compare my discussion of Michael Smith's distinction between the "deliberative" and "intentional" perspective above in footnote 20 to Chapter 3.

[148] Nagel 1986. Clearly the distinction developed here owes a more general debt to Nagel's work, especially Nagel 1974. Nagel himself addresses the problem of freedom in Chapter 7 of Nagel 1986, and there invokes something like this distinction between the first- and third-personal perspectives. Otherwise the view is quite different from the one developed here. In particular, Nagel factors the problem of freedom into what he calls the problem of "autonomy" and the problem of "responsibility," neither of which quite aligns with the problem with which I am concerned here, namely the nature of the freedom relation. This makes it difficult to evaluate Nagel's provocative last word on freedom: "Nothing approaching the truth has yet be said on the subject" (p. 137).

[149] What about my evidence that you are in pain: first-personal or not? I say that it is first-personal as well. But this would seem to rule out the scientific study of pain. Here we must be careful. On one natural view, what physiologists and other pain-investigators discover is the physical correlates of pain. But this is not yet evidence of pain. For this would entail that a character who has lived all her life without pain-sensations yet who has studied pain science extensively (in the manner of Jackson 1982) has evidence that there is pain. I am inclined to think that this body of evidence is evidence that there is *pain* only for someone who is already first-personally acquainted with pain.

[150] This principle does not rule out the possibility of going in the *opposite* direction. That is, we might accumulate third-personal evidence that determinism is true and, running the consequence argument as a *modus tollens*, conclude that the freedom relation is never instantiated. This seems antithetical to the spirit of compatibilism, but nothing I have said here rule it out. A more complete defense of compatibilism would be concerned not only with the limits of the first-personal perspective, which has been my concern here, but also with its seeming resilience in the face of certain sort of challenges derived from third-personal evidence.

(CP*), unlike some of our earlier forays, seems to successfully include the evidence that the compatibilist wishes to include. Scientific evidence ignored by actual contemporary scientists may still be third-personal in the relevant sense, and so included by (CP*). And *a priori* evidence, such as that obtained through thought experiments, is included as well. So, provisionally, (CP*) does not seem to wrongly exclude any of the sorts of evidence which the compatibilist grants might tell against the truth or falsity of determinism. The question then is whether it rightly *excludes* the evidence which the compatibilist thinks does *not* tell against the truth or falsity of determinism, in particular our evidence that the freedom relation is instantiated.

It excludes this evidence just in case all of our evidence that the freedom relation is instantiated is first-personal. Whether this is so will depend on our answer to a question that we have been pursuing throughout, namely what justification we have for the instantiation of the freedom relation.

If the account defended in Chapter 4 is correct, then our evidence for the instantiation of the freedom relation is first-personal. For our evidence is our immediate awareness of our own extended selves, a perspective not available to someone who occupies the view from nowhere. This is the sense in which the account vindicates the intuition voiced by Christine Korsgaard: "freedom . . . is not a theoretical property which can also be seen by scientists considering the agent . . . third-personally and from outside." [151] Our evidence that it is instantiated, in short, is precisely the sort of evidence that is excluded by (CP*).

It is less clear whether, if I am *wrong* about the nature of our evidence that the freedom relation is instantiated, whether I am nonetheless *right* that (CP*) excludes the evidence we do have. But I am inclined to think that this exclusion does succeed even on other accounts of our justification for believing that the freedom relation is instantiated. In particular, those who ground the epistemology of freedom in moral responsibility are appealing to evidence that seems to be paradigmatically first-personal, the sort of thing accessible only by someone who occupies a particular viewpoint on the world.

Indeed, this is a leitmotif of Strawson's "Freedom and Resentment":

> The vital thing can be restored by attending to that complicated web of attitudes and feelings which form an essential part of the moral life as we know it, and which are quite opposed to objectivity of attitude. Only by attending to this range of attitudes can we recover from the facts as we know them a sense of what we mean, i.e. of *all* we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice.[152]

Whether or not we agree with the bulk of Strawson's conclusions, there is much to be said for the thought that – in some sense or other – questions of responsibility arise only

[151] Korsgaard 1996. I quoted this remark above and in full in Chapter 3.1. Korsgaard herself would probably dissent from the positive account defended here, for on her view it is the process of deliberation rather than the faculty of bodily awareness that is responsible for the apprehension of freedom.
[152] Strawson 1961, p. 91.

from within the point of view of a particular agent with particular sentiments, and that they are "opposed to objectivity of attitude," or to what I have been calling the third-personal perspective.

To defend this thought we would have to inquire much more carefully into the epistemology of moral responsibility, as well as into the admittedly elusive boundary between the first- and third-personal perspectives. These inquiries seem to me very much worth pursuing, but from the point of view of this essay they are moot. For I have argued that our evidence for the instantiation of the freedom relation arises from a rather different source, namely from our perception of ourselves. And this sort of evidence is paradigmatically first-personal. If I am correct about this, then (CP*) does indeed exclude our evidence for the instantiation of the freedom relation from the body of evidence that counts for or against the truth of determinism.

Let us review. Our question was why the compatibilist was antecedently convinced that the consequence argument was unsound; that is, why he was a compatibilist. We answered that the compatibilist accepted a certain principle about what sort of considerations could be evidence for or against the truth of determinism. In particular, he holds that *first-personal* evidence is never evidence for or against determinism; this was the claim of (CP*). And since, according to the account I defended in Chapter 4, our evidence for the instantiation of the freedom relation is first-personal, the compatibilist holds, by (CP*), that evidence is not evidence for or against the truth of determinism. That is why he is antecedently convinced that the consequence argument, or *any* attempt to demonstrate that evidence for the instantiation of the freedom relation is evidence against the truth of determinism, is unsound. That is why he in turn thinks that there must be a flaw in any such argument or, if no such flaw can be found, why we should revise our ways of speaking in order to avoid the argument's conclusion.[153]

But why should we accept (CP*)? This question may have no informative answer; I, at least, am not able to provide one here. I myself find it immensely plausible: it seems to express in a concise way the sort of considerations that can, and cannot, be taken as properly scientific evidence. Saying more about (CP*) would require at least pursuing questions well beyond the topic of freedom, and hence outside the ken of this essay. For related principles seem also to apply in other debates. For example, in the debate between the dualist and the materialist about consciousness, the materialist's objection seems centrally motivated by the following thought:

(MP) Only *third-personal* evidence is evidence for the truth or falsity of physicalism

[153] This is not to say that the compatibilist takes himself to have first-personal evidence as to what the flaw is in the consequence argument, or even that there is some flaw or other in it. His first-personal evidence is only for the claim that the freedom relation is instantiated. His evidence that there is a flaw in the argument derives from his justification for his broader philosophical convictions, whatever its nature may be.

The dualist's arguments clearly violate this principle, for they proceed via *a priori* steps from the immediate data of experience to the conclusion that physicalism is false.[154]

It would be interesting to further pursue these parallels between the materialist and the compatibilist, and between the epistemic principles that they advocate.[155] For now our question is simply to say what *motivates* the compatibilist position in the first place. This we have done: what motivates compatibilism is the thought that (CP*) expresses a highly plausible epistemic principle, one that is violated by the incompatibilist's arguments.

5.8 The consequence argument and "transmission failure"

In Chapter 5.6 I claimed that the soundness of the consequence argument, in the more precise version proposed by van Inwagen, turns on the truth of (FS), a proposition whose truth I do not know how to determine. I therefore proposed a disjunctive strategy. If (FS) is false, then the Lewisian response goes through as a successful *hermeneutic* account of why our talk about freedom does not commit us to being free to break the laws in a deterministic world. If (FS) is true, matters are more complicated. I suggested that in that case the Lewisian response should be taken as a *revolutionary* proposal, a proposal to change our way of speaking about freedom. As I pointed out there, this proposal may face formal difficulties: in particular, it is not clear that the *subtraction* required for this proposal to succeed yields a determinate solution.

What is more worrisome, perhaps, is the marked diminishment of the compatibilist ambition involved in this proposal. For the compatibilist has effectively *granted* the incompatibilist's argument; his only recourse is to say that we may change our ways of speaking so as to avoid that argument's implications. For those reasons, I suggested that it would nice to have a better view for the compatibilist to take, if (FS) is indeed true. In this section I will offer just such a view.

This view begins with the observation, noted earlier, that a certain *epistemic* principle is needed for the incompatibilist's arguments to go through. This principle, recall, is:

---

[154] Notably in Chalmers 1996. The materialist's rejection of these arguments seems to rest especially heavily on (MP) when he also rejects any reduction of conscious experience, and so accepts what Chalmers calls "Type-B materialism." As Chalmers argues to great effect, this position can seem *ad hoc*: it is difficult to see what motivates it beyond its preservation of the truth of (MP). But, as in the case of compatibilism, this may well be motivation enough. There is a clear parallel between the sort of non-reductive but compatibilist view of freedom taken in this essay and the non-reductive but materialist view of consciousness taken by the "Type-B Materialist," and I suspect that a fully adequate defense of either one of these views might also yield a defense of the other.

[155] There may also be some significant disanalogies. For one thing, determinism is a scientific hypothesis while physicalism is sometimes said to be not a scientific but a "metaphysical" hypothesis. It is difficult to know, however, what this difference comes to. It is tempting to define this difference in terms of the different sorts of evidence that might be brought to bear for or against it, but if that is correct then the question of whether or not physicalism is a scientific or a metaphysical hypothesis may itself be a substantive one, for it is precisely these sorts of evidential principles, like (MP), about which the materialist and the dualist disagree.

(APT) If E is evidence that p, and it is a priori that if p then q, then E is evidence that q *just in virtue of* being evidence that p

The compatibilist view that I will defend here will turn crucially on the *denial* of (APT). On this view, while evidence *normally* transmits in the way that (APT) says it does, it does not *always* do so, and in particular it fails to transmit in the consequence argument.

Let us begin by saying why some think that (APT) is not *always* true. Their reasons for doing so have nothing to do with the consequence argument. They arise rather from cases like the following.[156] Imagine someone has perfectly good evidence E that the bird before him is a canary: evidence of the ordinary sort, such as looking at the bird before him, listening to its song, consulting his guide to birds, and so forth. Then consider the following four propositions:

(1) E is evidence that this bird is a canary
(2) It is a priori that, if this bird is a canary, then it is not a cleverly disguised raven
(3) E is *not* evidence that this bird is not a cleverly disguised raven just in virtue of being evidence that this bird is a canary
(4) (APT)

These four propositions cannot be true together. Which should we reject? A natural line of reasoning may lead someone to reject (4). We should not reject (2), for it *is* a priori that nothing can be a canary and a raven. Nor should we reject (1), for if there is ever ordinary evidence that a bird is a canary, as there surely is, then there is in this case. But neither it seems, should we reject (3). For the evidence gathered from this sort of ordinary observation and research is no evidence at all that this bird is not a cleverly disguised raven, for (provided the disguise is clever enough) that evidence does not put one in a position to *discriminate* between such a bird and an ordinary canary. And if each of those claims is correct, then (4) – that is, (APT) – *must* be false.

Cases like this, cases where the principle that I am calling (APT), fail to hold, have come to be called cases of "transmission failure."[157] It is a matter of controversy when and why transmission failure occurs, and indeed even whether there *is* such a thing.[158] These questions seem to me very much open, and I will not say anything to resolve them here. The role such cases have to play here is rather delimited. They serve to prepare the way for a novel response to the consequence argument – one that involves the denial of (APT).

The response is simply this. It is to grant that, if (FS) is indeed true, that each of the premises of the consequence argument that make claims about the freedom relation and the laws are true. The false move of the consequence argument is the crucial epistemic one, namely the move from:

---

[156] For cases of this form, though not the present diagnosis of them, see Dretske 1970.
[157] Notably by Crispin Wright. See, for example, Wright 2000.
[158] For one recent set of criticisms, see Silins 2005.

(1) E is evidence that S was free to A . . .
(9) It is a priori that, if S was free to A, then determinism is false

to:

(C) E is evidence that determinism is false just in virtue of being evidence that S was free to A (by 1, 9 and (APT))

This move fails, says the compatibilist, because (APT) fails in the present case. So while the consequence argument is sound from the ontological point of view, it fails for epistemic reasons.

This move can seem deeply *ad hoc*. But the compatibilist has a defense for it. He claims that, *first*, we have independent grounds, ones having nothing to do with the consequence argument, for denying (APT). These grounds arise instead from the more mundane cases of "transmission failure" surveyed above.[159] And he claims that, *second*, (CP*) and principles like it (such as (MP)) express deep and plausible constraints on the limitations of our epistemic situation. Indeed, they are *so* deep and *so* plausible, says the compatibilist, that when we face a conflict between (CP*) and the generally but not universally applicable epistemic principle (APT), as we do in the present case, it is (APT) that we should give up.

If this response is to be taken seriously, then the compatibilist owes us a much more extended defense of (CP*) than I have been able to offer here. Indeed, he owes us a good bit more than that. For the most natural development of the present idea involves a quite general limit on the power of *a priori* argumentation which proceeds from first-personal evidence (a constraint that applies also, as I have said, to the dualist's arguments). The failure of the consequence argument is, from this point of view, but one way in which that limit makes itself known.

5.9 Epistemic compatibilism defended

We are now in a position to say what compatibilism is and how it is to be defended. The best version of compatibilism defends an *epistemic* principle, one which *denies* that there is an essential connection between evidence for the freedom relation and evidence for determinism. The best argument against this principle – that is, for incompatibilism – is the consequence argument. How the compatibilist is to respond to this argument depends on the truth of the contentious claim (FS). If (FS) is false, then the compatibilist ought to adopt David Lewis's diagnosis of where the consequence argument goes wrong. If (FS) is true, the compatibilist ought to either accept Lewis's proposal as a *revolutionary*

[159] There is an important disanalogy between the consequence argument and the arguments that are exemplars of transmission failure. This is that we typically accept the conclusions of the latter arguments (accept, for example, that the bird is not a cleverly-disguised raven) while the compatibilist we are considering here denies or at least suspends judgment on the claim that determinism is false. But it is not clear that this disanalogy undermines the compatibilist's argument. For may grant the point and yet insist, first, that the standard cases of transmission failure do at least show that (APT) sometimes fails and that, second, once this is granted, it opens the possibility of denying (APT) in other cases, including cases that are in important respects quite different from standard cases of transmission failure.

proposal about our talk about freedom, or he ought to simply deny the epistemic principle that is needed for the consequence argument to go through, namely (APT). As I have suggested, it is the latter strategy which seems to me the better one. For it promises to achieve nothing less that the compatibilist ambition, which is to say that freedom as we actually speak and think about it is no evidence at all for the falsehood of determinism.

It may be objected that the last proposal limits the compatibilist ambition in a rather different way. For we would expect from the compatibilist an answer to the question of whether freedom *really is* compatible with determinism. But the last defense of compatibilism offers no such answer. I am inclined to think that the request for such an answer demands too much. The best we can expect of a defense of a position like compatibilism is a principled account of how we ought to *apportion our credences*, given our imperfect epistemic situation. And epistemic compatibilism does precisely this.[160]

This position may appear all too close to a disappointing quietism about the metaphysical. But I think this appearance is misleading. For even once we accept this limit on the ambitions of compatibilism, there remain difficult metaphysical questions in the area, ones to which the compatibilist still owes us an answer if his position is to be a credible one. Let me mention one that seems to me particularly difficult.

This problem is best stated in terms of the *supervenience* relation.[161] Let us say that some set of facts A *supervenes on* some set of facts B just in case there is a difference in the A-facts *only if* there is a difference in the B-facts. That is, it is not possible that the B-facts are the same in two possible worlds but the A-facts in those worlds are different. Now we can state the *supervenience question*:

(SQ) Do the freedom facts supervene on the non-freedom facts?

That is, do the set of facts about the freedom relation (and what we have earlier called *trivial transformations* of the freedom relation) supervene on, in the sense of supervenience just given, supervene on all the *other* facts, that is on the complement of that set?

If our answer is *no*, then we face a rather interesting set of difficulties. Most clearly, this answer sits very uneasily with what I take to be our intuitions about the freedom relation. According to this view, it is possible that there are two worlds alike in respect to all of

---

[160] Another way of casting doubt on this proposal arises simply from making explicit the dialectical situation its proponent finds himself in. He accepts that the freedom relation is instantiated, he accepts that an argument from the instantiation of freedom to the falsehood of determinism is sound, and still he denies the truth of determinism. This position is, we might say, dialectically unsatisfying; a conversation on the topic of freedom and determinism with the proponent of such a strategy promises to be frustrating for his interlocutor. But it remains unclear why this is sufficient to show that the proponent of such a strategy holds beliefs that are *irrational*.

[161] See Kim 1984 for a useful discussion of this relation. Note that throughout that what I refer to as supervenience is a relation that holds *necessarily* between two sets of facts, rather than merely relative to the actual laws or to some other constraint. It is, in short, what other authors refer to as "metaphysical supervenience."

their physical properties, their experiential properties, and so forth, but which are different in the facts about what agents are free to do. To make this difference still more striking, we may imagine *freedom zombies*, agents who are *exactly* like us in all respects but for the fact that they are not free to perform any actions but for those that they actually perform.[162] All of this is, admittedly, somewhat strange.

Nonetheless, I think an outright rejection of this possibility rests too much on the variable data of intuition. For one thing, our intuitions in this case do not seem quite as incontestable as in the case I will consider presently, namely the supervenience of moral facts on non-moral facts. This consideration does not, admittedly, outweigh the intuitive oddity of a negative answer to (SQ). But it seems to me worth thinking through this answer, if only because it represents a heretofore unexplored position on a topic which is sometimes alleged to be lacking in novel points of view.[163]

The difficulty I alluded to earlier, however, is one that arises when we give what I take to be the more plausible answer to (SQ), and accept that the freedom facts *do* supervene on the non-freedom facts. Here we are faced with a familiar problem for a non-reductive view like mine, one pressed most forcefully by Simon Blackburn in the case of *moral* facts.[164] For we are granting that facts about freedom (or, in Blackburn's case, moral facts) supervene on non-freedom (or non-moral) facts. But this seems like a fact that needs to be *explained*: why can these facts *not* "come apart" from each other? The reductivist has a simple answer to this question: facts about freedom or moral facts *supervene on* other facts just because they *reduce to* those facts. But the non-reductivist seems to have no tenable answer to this question. This challenge has seemed to many to support the thought that there *are* no moral facts (this is Blackburn's conclusion). And it might equivalently be thought to support the conclusion that there are no freedom facts.[165]

---

[162] So named by analogy with the more familiar zombies of the consciousness literature. See Chalmers 1996, pp. 94-99.

[163] Note that the analogy I have been pressing between compatibilism about freedom and materialism about the mind begins to strain here. For this view – that is the conjunction of the view I have been defending conjoined with the denial of the supervenience of freedom – is compatibilist in that it accepts (ECOMP**) as well as (CP*). But in some respects this view is quite close to the dualism of the sort developed in Chalmers 1996. So the parallel between these debates grows somewhat less straightforward once we begin to understand the *range* of views that are available on the question of freedom.

[164] See Blackburn 1985.

[165] Is not this phenomenon much more general, and therefore perhaps less worrisome? After all, we cannot give a reductive definition of what it is to be (for example) a table, and yet it is reasonable to believe that (a) there are table-facts that (b) these supervene on non-table facts. So why we should be especially worried about the normative and about freedom? The first thing to say is that showing that a problem is general does not make that problem go away; this may also be an argument for rejecting (a) or (b). But let us grant that this argument fails in the case of tables. Would not this show that it must fail in the case of freedom as well? Not necessarily. For one thing, the obstacles to a reductive analysis in the case of tablehood arise from different sources; they seem to be entangled in some way with the phenomenon of vagueness, with the claim that not every object is either clearly a table or clearly not a table. The primitivism of freedom and of the normative arise from a different and, we might say, more fundamental irreducibility – nowhere in the foregoing arguments have I appealed to the vagueness of freedom (indeed I *conjecture* that freedom is not vague, so that it is always the case that S is clearly free to A or S is clearly not free to A). So this marks at least one disanalogy: while we might somehow reject the supervenience argument in the case of

As I say, this seems to me a difficult problem, and I have no answer to it. I think that freedom facts *do* supervene on freedom facts, yet I have no explanation of why this is so. Furthermore, it seems that such an explanation is a desideratum for an adequate theory of freedom, and it seems to me that the primitivist view I favor is in no position to offer such an explanation. I do not on these grounds reject the present view of freedom, because it seems more credible to me that there are inexplicable superveniences than it is that the freedom relation is never instantiated. But I grant that there is something unsatisfactory about my view here, and that it would be intellectually preferable to offer a better answer to this question. This is one way in which the present view is not a complete theory of the freedom relation, but only the beginning of one. And it makes clear, I hope, that there is further metaphysical inquiry to be done here, even if we accept that, I have suggested, that there are in the end limits to what such inquiry can reveal.

5.10 Reduction revisited

The conclusion of Chapter 2, recall, was that the freedom relation was simple and unanalyzable. My argument there proceeded by a kind of induction on failed analyses. This argument was a good one, but we are now in a position to state a deeper objection to the reduction of the freedom relation. This is that such a reduction is profoundly *unmotivated*. For if the goal of offering a reduction was to defend compatibilism – by satisfying what I called the *dialectical* and the *ontological* roles of a reduction – then it turns out that goal can be met by other means. For I have tried to show here how the compatibilist can answer the incompatibilist's arguments, as well as defending the compatibility of freedom and determinism to his own satisfaction, without attempting to reduce the freedom relation to something allegedly more basic.

My hope is that the present chapter has thereby made both compatibilism *and* primitivism more plausible. Moreover, I think it has put us in a position to diagnosis at least *one* source of the familiar intractability of the debate over compatibilism. For from the present point of view the standard views on freedom have consisted of a conjunction of an implausible view about the freedom relation (reductivism) conjoined with a plausible view of what it is evidence for (compatibilism), opposed to a plausible view of the freedom relation (primitivism) conjoined with an implausible view of what it is evidence for (incompatibilism). But, if the present line of argument is sound, then these combinations are not mandatory. We may take a view of the freedom relation that is primitivist *and* compatibilist. That is in fact, as I have argued here, the view that we *ought* to take.

---

tables by gesturing towards the problematic phenomenon of vagueness, no such appeal will help us in the case of freedom.

The Possibility of Freedom

Conclusion

Near the outset of this essay, I distinguished three questions about the metaphysics of the freedom relation:

(i) Is the freedom relation ever instantiated?
(ii) Is the freedom relation *reducible* to some other relation?
(iii) Does the instantiation of the relation *entail* the falsehood of determinism?

And three questions about its epistemology:

(iv) Do we have any *evidence* that the freedom relation is instantiated?
(v) Is the instantiation of the freedom relation *inferable* from some other facts?
(vi) Does our evidence for the instantiation of the freedom relation *justify* the belief that determinism is false?

I have now offered answers to four of these questions. I gave negative answers to questions (ii) and (v): the freedom relation is neither reducible to nor inferable from any other relation. There is, I said, a *does-can* gap between facts about freedom and all other facts. And I defended a positive answer to question (iv): we have, I argued, *perceptual* evidence that the freedom relation is instantiated. Finally, I argued, this evidence does not give us justification for the belief that determinism is false. So we should also give a negative answer to question (vi), and should therefore be *compatibilists*, in at least one sense of that term.

I argued for each of these answers on its own merits, but taken together they cohere into something like the beginning of a *theory* of the freedom relation. On this view, freedom is a simple fact, one that is something "over and above" the world as it is revealed by the sciences, and for which our justification is immediate. This view is akin to certain versions of "intuitionism" about the ethical, but for the fact that it does not need to postulate any faculty of "insight" to explain how we know about freedom. Rather, on this view, freedom is something revealed to us through our ordinary faculties of bodily awareness.

Let me close by addressing the two questions that I have not yet answered, namely questions (i) and (iii).

Question (iii) I discussed in the previous chapter. There I did not claim either that the instantiation of the freedom relation did, or that it did not, entail the falsehood of determinism. I did, however, make something like a meta-claim about this question. This is that its answer is not crucial to determining the truth of compatibilism, on the best formulation of compatibilism. Our answer to it will rather dictate the *form* that compatibilism should take. If the answer to (iii) is negative, then the compatibilist ought

to offer a diagnosis of where the incompatibilist's arguments go wrong very much like that defended by David Lewis. If the answer to (iii) is positive, then the compatibilist ought to offer what I called a "revolutionary" version of Lewis's proposal, or claim that a kind of "transmission-failure" prevents our evidence for the instantiation of the freedom relation from being evidence for the falsehood of determinism.

It is an important question which of these ways we should take, but it is downstream from the question of whether compatibilism is true. *That* question is to be decided on much more general epistemic grounds, such as the relationship between first-personal and third-personal evidence. These grounds are not particular to freedom. They extend quickly into other debates, such as that over whether consciousness requires the falsehood of physicalism, debates which are beyond the ken of this essay. In this sense, I have not offered a complete defense of compatibilism here, but have rather set forth the general way in which I think compatibilism ought to be defended.

This leaves question (i), which may fairly be regarded as the first question in the metaphysics of freedom. This question is one that I will not answer here. I *have* claimed that we do have evidence for the instantiation of the freedom relation. This of course bears on question (i), but it does not decide it. This is partly for the very general reason that our evidence may always be misleading, and so may be misleading in the present case. But there are three somewhat more specific considerations that recommend a certain modesty in our metaphysical speculations about freedom. Let me close by reviewing them.

The first consideration is the idea that temporary or permanent limits on our epistemic situation render us *incapable* of knowing the answer to question (i). This position we might call *mysterianism* about freedom. The mysterian position is motivated by the thought that the answers to certain intractable philosophical questions may simply lie beyond our ken.[166] I say that this consideration applies specifically to the case of freedom precisely because question (i) is so famously intractable. There is admittedly something unsatisfactory about the mysterian position, and it no doubt constitutes a last resort. Nonetheless, I think that we ought to at least take seriously the thought that the answer to question (i) is simply, for us, unknowable.

The second consideration is one indicated briefly at the close of Chapter 1. This is that the deliverances of the sciences, and especially of experimental psychology, may bear on the question of whether agents are free to perform actions, even if it is only with great care that we can say exactly *how* they bear on this question. Experiments purporting to reveal "the springs of action" have been of great interest in recent psychology and philosophy and, while I think we should not be too credulous about the claims made by the proponents of these experiments, neither ought we claim (as do proponents of "paradigm case" arguments about freedom) that such experiments could *never* bear on the truth of our ordinary claims about freedom. So this is another consideration that ought to properly constrain our speculations about question (i): that there are empirical investigations that may bear on it, and that these investigations remain open.

---

[166] This view is defended in McGinn 1993; for this sort of view of freedom, see pp. 79-92.

The third consideration arises from limits of this essay that I have already noted. A proper evaluation of the question of compatibilism, I argued, requires that we look carefully not only at freedom but also at other properties that seem to stand in some tension with the natural world, such as consciousness and value. So too, I think, does a proper evaluation of question (i). The metaphysics suggested by the present essay is one in which freedom is irreducible and in which certain aspects of its nature, such as its supervenience on the physical, are left unexplained. This approach is a fragmentary one, and it should make us hesitant about answering question (i) in the affirmative. For to properly answer this question, we need to try to answer similar questions about consciousness, and responsiveness to value, and the other aspects of animate beings that distinguish them from mere things like tables and rocks. It is only from this more general point of view, I think, that we can defend an answer to question (i) in which we can be reasonably confident. It may be, of course, that a patchwork metaphysics is the best that we can do. But the question of whether we can do better is not one that should be decided hastily.

The Possibility of Freedom

Bibliography

Albritton, Rogers. 1985. "Freedom of Will and Freedom of Action." In Watson 2003.

Alston, William. 2005. "Perception and Representation." *Philosophy and Phenomenological Research* 70: 253-289.

Austin, J.L. 1956. "Ifs and Cans." *Philosophical Papers* (1961). Oxford UP.

Blackburn, Simon. 1985. "Supervenience Revisited." *Essays in Quasi-Realism* (1993). Oxford UP.

Bok, Hillary. 1998. *Freedom and Responsibility*. Princeton UP.

Bobzien, Suzanne. 1998. *Determinism and Freedom in Stoic Philosophy*. Oxford UP.

Bonevac, Daniel et al. 2006. "The Conditional Fallacy." *The Philosophical Review* 115: 273-316.

Broome, John. 1999. "Normative Requirements." *Ratio* 12: 398-419.

Byrne, Alex and Hilbert, David (eds.). 1997. *Readings on Color, Volume 1: The Philosophy of Color*. MIT Press.

Burgess, John. 1983. "Why I Am Not a Nominalist." *Notre Dame Journal of Formal Logic* 24: 93-105.

Campbell, John. 1993. "A Simple View of Colour." In Byrne and Hilbert 1997.

Chalmers, David. 1996. *The Conscious Mind*. Oxford UP.

Chisholm, Roderick. 1976. "Human Freedom and the Self." In Watson 2003.

Danto, Arthur. 1963. "What We Can Do." *The Journal of Philosophy* 60: 435-445.

Danto, Arthur. 1965. "Basic Actions." *American Philosophical Quarterly* 2: 141-148.

Davidson, Donald. 1963. "Actions, Reasons, Causes." In Davidson 1980.

Davidson, Donald. 1970. "How Is Weakness of the Will Possible?" In Davidson 1980.

Davidson, Donald. 1971. "Agency." In Davidson 1980.

Davidson, Donald. 1973. "Freedom to Act." In Davidson 1980.

Davidson, Donald. 1980. *Essays on Actions and Events*. Oxford UP.

Dretske, Fred. 1970. "Epistemic Operators." *The Journal of Philosophy* 67: 1007-1023.

Earman, John. 1986. *A Primer on Determinism*. D. Reidel.

Fara, Michael. 2001. *Dispositions and their Ascriptions*. Dissertation, Princeton.

Fischer, John Martin. 2002. "Frankfurt-Style Compatibilism." *My Way: Essays on Moral Responsibility* (2006). Oxford UP.

Fischer, John Martin and Ravizza, Mark. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge UP.

Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility." In Frankfurt 1988.

Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of a Person." In Frankfurt 1988.

Frankfurt, Harry. 1987. "Identification and Wholeheartedness." In Frankfurt 1988.

Frankfurt, Harry. 1988. *The Importance of What We Care About*. Cambridge UP.

Gibson, J.J. 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin.

Grice, H.P. 1961. "The Causal Theory of Perception." *Proceedings of the Aristotelian Society* 121: 121-152.

Hampshire, Stuart. 1975. *Freedom of the Individual*. Princeton UP.

Harman, Gilbert. 1965. "The Inference to the Best Explanation." *The Philosophical Review* 74: 88-95.

Harman, Gilbert. 1977. *The Nature of Morality*. Oxford UP.

Harman, Gilbert. 1990. "The Intrinsic Quality of Experience." *Reason, Meaning, and Mind* (1999). Oxford UP.

Hornsby, Jennifer. 1980. *Actions*. Routledge & Kegan Paul.

Huemer, Michael. 2005. *Ethical Intuitionism*. Palgrave Macmillan.

Hume, David. 1739. *A Treatise of Human Nature*. L.A. Selby-Bigge, ed. (1978). Oxford UP.

Hume, David. 1748. *An Enquiry Concerning Human Understanding*. Tom Beauchamp, ed. (1999). Oxford UP.

Hursthouse, Rosalind. 1991. "Arational Action." *The Journal of Philosophy* 88: 57-68.

Jackson, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32: 127-136.

Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford UP.

James, William. 1884. "The Dilemma of Determinism." *The Will to Believe and Other Essays in Popular Philosophy* (1956). Dover.

Johnston, Mark. 1992. "How to Speak of the Colors." In Byrne and Hilbert 1997.

Kelly, Thomas. 2005. "Moorean Facts and Belief Revision, or, Can the Skeptic Win?" *Philosophical Perspectives* 19: 179-209.

Kelly, Thomas. 2006. "Evidence." *Stanford Encyclopedia of Philosophy*.

Kim, Jaegwon. 1984. "Concepts of Supervenience." *Philosophy and Phenomenological Research* 45: 153-176.

Korsgaard, Christine. 1996. *The Sources of Normativity*. Cambridge UP.

Kripke, Saul. 1980. *Naming and Necessity*. Harvard UP.

Lehrer, Keith. 1960. "Can We Know that We Have Free Will by Introspection?" *The Journal of Philosophy* 57: 145-156.

Lehrer, Keith. 1966. "An Empirical Disproof of Determinism?" Lehrer, ed.. *Freedom and Determinism*. Random House.

Lehrer, Keith. 1968. "Cans without Ifs." *Analysis* 29: 29-32.

Lewis, David. 1973. *Counterfactuals*. Blackwell.

Lewis, David. 1981. "Are We Free to Break the Laws?" *Philosophical Papers, Volume 2* (1987). Oxford UP.

Lewis, David. 1983. "New Work for a Theory of Universals." In Lewis 1999.

Lewis, David. 1998. "Finkish Dispositions." In Lewis 1999.

Lewis, David. 1999. *Papers in Metaphysics and Epistemology*. Cambridge UP.

Lipton, Peter. 2004. *Inference to the Best Explanation, 2nd Edition*. Routledge.

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

Manley, David and Wasserman, Ryan. 2008. "On Linking Dispositions and Conditionals." *Mind* 117: 59-84.

Martin, C.B. 1996. "Dispositions and Conditionals." *The Philosophical Quarterly* 44:1-8.

McGinn, Colin. 1993. *Problems in Philosophy: The Limits of Inquiry*. Blackwell.

Mele, Alfred. 2002. "Agents' Abilities." *Nous* 37: 447-470.

Mele, Alfred. 2006. *Free Will and Luck*. Oxford UP.

Mill, J.S. 1865. *An Examination of Sir William Hamilton's Philosophy*. J.M. Robson, ed. (1963). U of Toronto P.

Moore, G.E. 1903. *Principia Ethica*. Cambridge UP.

Nagel, Thomas. 1974. "What Is It Like To Be a Bat?" *The Philosophical Review* 83: 435-450.

Nagel, Thomas. 1986. *The View From Nowhere*. Oxford UP.

Nahmias, Eddy et al. 2006. "Is Incompatibilism Intuitive?" *Philosophy and Phenomenological Research* 73: 28-53.

O'Connor, Timothy. 2000. *Persons and Causes*. Oxford UP.

Peacocke, Christopher. 1979. *Holistic Explanation: Action, Space, Interpretation*. Oxford UP.

Peacocke, Christopher. 1999. *Being Known*. Oxford UP.

Prior, Elizabeth et al. 1982. "Three Theses about Dispositions." *American Philosophical Quarterly* 19: 251-257.

Pryor, James. 2000. "The Skeptic and the Dogmatist." *Nous* 34: 517-549.

Pryor, James. 2005. "There Is Immediate Justification." Matthias Steup and Ernest Sosa, eds. *Contemporary Debates in Epistemology*. Blackwell.

Rawls, John. 1971. *A Theory of Justice*. Harvard UP.

Reid, Thomas. 1788. *Essays on the Active Powers of Man*. Keith Lehrer and Ronald Beanblossom, eds. (1975). Bobbs-Merrill.

Rosen, Gideon. 2004. "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18: 295-313.

Siegel, Susanna. 2005. "The Phenomenology of Efficacy." *Philosophical Topics* 33:1.

Silins, Nico. 2005. "Transmission Failure Failure." *Philosophical Studies* 126: 71-102.

Smith, Michael. 1994. *The Moral Problem*. Blackwell.

Smith, Michael. 1998. "The Possibility of Philosophy of Action." *Ethics and the A Priori* (2004). Cambridge UP.

Strawson, P.F. 1961. "Freedom and Resentment." In Watson 2003.

Taylor, Richard. 1960. "I Can." *The Philosophical Review* 69: 78-89.

Taylor, Richard. 1966. *Action and Purpose*. Prentice-Hall.

van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford UP.

van Inwagen, Peter. 2000. "Free Will Remains a Mystery." *Philosophical Perspectives* 14: 1-19.

van Inwagen, Peter. 2004. "Freedom to Break the Laws." *Midwest Studies in Philosophy* 28: 334-350.

Velleman, David. 1989. "Epistemic Freedom." In Velleman 2000.

Velleman, David. 1992. "What Happens When Someone Acts?" In Velleman 2000.

Velleman, David. 2000. *The Possibility of Practical Reason*. Oxford UP.

Vivhelin, Kadri. 2004. "Free Will Demystified: A Dispositionalist Account." *Philosophical Topics* 32: 427-450.

Vivhelin, Kadri. 2007. "Compatibilism, Incompatibilism, and Impossibilism." Theodore Sider et al., eds. *Contemporary Debates in Metaphysics*. Blackwell.

Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Harvard UP.

Watson, Gary (ed.). 2003. *Free Will, 2nd Edition*. Oxford UP.

Wegner, Daniel. 2002. *The Illusion of Conscious Will*. MIT Press.

White, Roger. 2006. “Problems for Dogmatism.” *Philosophical Studies* 131: 525-557.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford UP.

Wolf, Susan. 1990. *Freedom within Reason*. Oxford UP.

Wright, Crispin. 2002. “(Anti)-Skeptics Simple and Subtle: Moore and McDowell.” *Philosophy and Phenomenological Research* 65: 330-348.

Wright, Crispin. 2004. “Warrant for Nothing (and Foundations for Free)?” *Aristotelian Society Supplementary Volume* 78: 167-212.