

Does Consciousness Cause Behavior?

edited by Susan Pockett, William P. Banks, and Shaun Gallagher

**The MIT Press
Cambridge, Massachusetts
London, England**

- Velmans, M. 2003. Preconscious free will. *Journal of Consciousness Studies* 10 no. 12: 42–61.
- Wegner, D. 2002. *The Illusion of Conscious Will*. MIT Press.
- Wegner, D., and Wheatley, T. 1999. Apparent mental causation—sources of experience of will. *American Psychology* 54: 480–492.
- Williams, B. 1995. *Making Sense of Humanity*. Cambridge University Press.
- Zhu, J. 2003. Reclaiming volition: An alternative interpretation of Libet's experiments. *Journal of Consciousness Studies* 10, no. 11: 61–77.

7 Empirical Constraints on the Problem of Free Will

Peter W. Ross

With the success of cognitive science's interdisciplinary approach to studying the mind, many theorists have taken up the strategy of appealing to science to address long-standing disputes about metaphysics and the mind. For example, in the 1980s C. L. Hardin's *Color for Philosophers* introduced perceptual psychology into the discussion of the metaphysics of color. Psychological research, Hardin showed, can provide constraints for the philosophical debate, ruling out certain positions on the nature of color. To provide an example of such a constraint, psychophysics shows that for any determinate color, physical objects in the same viewing condition may differ in indefinitely many ways with respect to their spectrally relevant physical properties, and nevertheless look precisely that color. Neurophysiology explains how such physically distinct objects, called metamers, can occur. These scientific findings now provide constraints on philosophical accounts of color—anyone writing on the philosophy of color must acknowledge these scientific findings, which rule out, for instance, the proposal that colors are properties of physical objects which are physical natural kinds.¹

Recently, philosophers and psychologists have also begun to explore how science can be brought to bear on the debate about the problem of free will. Prominent examples have been Robert Kane's (1996, 1999) attempt to shore up libertarianism through an appeal to physics and neuroscience, Daniel Dennett's (2003) account of compatibilist freedom in evolutionary terms, and Daniel Wegner's (2002) psychological model of our experience of mental causation.

I take the problem of free will to be the problem of whether we control our actions.² (The problem of free will is best stated in terms of the first-person point of view. But in stating the problem this way I don't mean to endorse a necessary connection between consciousness and control, or to prejudge any theoretical positions at all.) Science has traditionally been central to the problem, for one standard way of viewing it is through considering whether we can fit ourselves as free agents into the natural world characterized by science. If our actions are brought about by a combination of genetic and environmental factors, we can fit ourselves into the

natural world. But in that case, since we ultimately control neither our genes nor our environments,³ it seems that we don't control our actions. The standard responses are that our actions are in fact brought about by genetic and environmental factors and this rules out control (hard determinism); that our actions are so brought about but nevertheless this doesn't rule out control (compatibilism); and that the bringing about of our actions involves a third factor, namely causal indeterminacy, which is necessary and in certain contexts sufficient for control (libertarianism).

Despite the centrality of science to the problem of free will, even Kane, Dennett, and Wegner do not consider the general question of how scientific research can provide constraints that serve to rule out certain positions on the problem; their use of empirical findings is tailored to support specific philosophical or psychological proposals. I will attempt to clarify the debate by taking up the general question of empirical constraints.

The debate can be viewed according to two basic dimensions: a dimension opposing compatibilism and incompatibilism, where the disagreement is about whether the bringing about of our actions by genetics and environment is compatible with control, and a dimension opposing indeterminism and determinism, where the dispute is about whether causal indeterminacy can ever be sufficient for control. I will argue that empirical findings don't apply at all to the dispute between compatibilism and incompatibilism. However, I will show that empirical research can provide constraints in connection with the other fundamental dimension, namely the dispute between libertarianism, which claims that causal indeterminacy is in certain contexts sufficient for controlling our actions, and the other positions, which deny this sufficiency claim. I will argue that psychological research into the accuracy of introspection has the potential to decide the truth of naturalized libertarianism, and thus that this research provides the source of the most powerful constraint, that is, the constraint that rules out the broadest category of positions.

1 Compatibilism versus Incompatibilism: A Semantic Debate

One of the fundamental issues associated with the problem of free will is whether the bringing about of our actions by genetic and environmental factors rules out control. A strong intuition supports the incompatibilist claim, embraced by hard determinists and libertarians, that it does. The intuition is that, despite our ordinary thought that we control at least some of our actions, if our actions in fact are produced by genetics and environment—factors ultimately outside of our control—then this ordinary thought is mistaken.

But of course a strong intuition also supports the ordinary thought that we control at least some of our actions. The compatibilist accepts that our actions are brought

about by genetics and environment and nevertheless attempts to hold on to the intuition of control. Consequently, the compatibilist seeks to counter the incompatibilist intuition by having us consider what we really mean by "control." If we mean *self-formation*, where some of our goal-directed states become causally efficacious in producing action in a way that is not the result of genetics and environment alone but is also the result of a third necessary factor (indeterminacy), then we don't have control in that sense.⁴ But, compatibilists argue, why think that by "control" we mean self-formation?

As an alternative, compatibilists give a variety of sophisticated accounts of the psychological complexity that allows us to take a standpoint from which—they contend—it makes sense to say that we control our actions despite their being brought about by genetics and environment. So, for example, Harry Frankfurt (1971) claims that free will is a matter of taking a standpoint from which we reflectively want certain desires to be effective in producing action and acting on those desires (rather than on desires which we reflectively don't want to be effective). Thus we control our actions in the sense that we take a reflective standpoint from which we choose our effective desires. Dennett's recent account also characterizes a reflective standpoint from which we choose our effective desires, a standpoint from which our choice is "rational, self-controlled, and not wildly misinformed" (2003, p. 284). In addition, Dennett attempts to explain the development of the standpoint so characterized in evolutionary terms.

Yet no matter how sophisticated and persuasive an account of a reflective standpoint may be, the incompatibilist contends that its perspective on desire and action produces mere illusion. We have the psychological complexity enabling us to take a standpoint from which it *seems* to make sense to say that we control our actions. But, the incompatibilist insists, if our actions are brought about genetics and environment—a point to which the compatibilist is committed—then they are out of our control.⁵

The debate between these positions amounts to whether we should accept that "control" means self-formation—which takes indeterminacy as a necessary factor—or that it has a meaning consistent with the bringing about of our actions by genetics and environment alone. But nothing is at stake here except what we mean by "control."

Although the debate between compatibilism and incompatibilism is semantic, it is not trivial. Both the compatibilist meaning of "control" and the incompatibilist meaning have intuitive pull. Yet the question of which of these meanings is fundamental is not one that we can address through the empirical sciences. (It is interesting to consider whether this debate could be settled at all, and if so, how. In view of my limited purposes here, I will set these questions aside.)

To illustrate the point that the empirical sciences can't address which meaning of "control" is fundamental, consider that the issue dividing compatibilists and hard determinists is merely incompatibilism (the issue of whether the bringing about of our actions by genetics and environment is compatible with control) and not any aspect of a scientific description of the world. Rather than disagree about a scientific description of the world, they disagree about how a scientific description of the world relates to our normative terms.

While the hard determinist understands an interrelated set of normative terms on the basis of self-formation, the compatibilist understands a homophonic set on the basis of a meaning of control without self-formation. As a result, they often talk past each other. However, because a hard determinist can accept that there is an ersatz control (that is, control without self-formation) from which we can derive similarly ersatz but nevertheless useful justification for punishment, this theorist is in a position to accept both sets of terms.⁶ As far as a hard determinist is concerned, compatibilism is correct in its account of ersatz control. Thus, a hard determinist can accept a compatibilist sense of "control" as practically indispensable, but retain a distinct sense of "control" which requires self-formation (and thus, by the hard determinist's lights is non-ersatz or genuine control); and, likewise, accept a compatibilist sense of "punishment" as justified in terms of its practical indispensability, but retain a distinct sense of "punishment" the justification of which requires self-formation (and thus, by the hard determinist's lights is genuine punishment); and so on.

Because these views do not differ on any aspect of a scientific description of the world with regard to ersatz control, and because they agree that we can't achieve self-formation, they don't differ on any aspect of a scientific description of the world at all. But then there is no way to address their dispute through the empirical sciences. Given any amount of empirical research, the compatibilist and the hard determinist will remain divided, flaunting different meanings of "control."⁷

Moreover, suggestions that compatibilists can refute hard determinism on non-semantic grounds are misguided, building more into hard determinism than an adherent of this view needs to accept. For instance, although Skinner's influential version of hard determinism was combined with a crude behaviorism, and the success of cognitive science has shown the empirical inadequacy of this behaviorism, hard determinism doesn't rely on Skinnerian behaviorism. Hard determinism merely claims that our actions are brought about by genetics and environment, and that this rules out control.

However, setting aside the dispute between hard determinism and compatibilism, whether libertarianism provides a tenable alternative is not a semantic issue. Libertarians are incompatibilists. But libertarians claim more than that what is meant by "control" is self-formation. They also contend that at least sometimes we *achieve*

self-formation, and thus that the conditions sufficient for self-formation exist. It is at this point that empirical research becomes relevant, at least if the metaphysics of self-formation is *naturalistic*, that is, scientifically tractable.

Thus, empirical findings apply only to the dimension of the debate pitting libertarianism, which claims that indeterminacy is in certain contexts sufficient for control, against the other positions, which deny this. In what follows, I will consider how scientific research can provide constraints for the debate between naturalized libertarianism and the so-called determinist views.

2 Quantum Indeterminacy: At Best a Weak Constraint

The general area of the problem of free will is, of course, standardly characterized as a problem having to do with determinism. And determinism, which claims that at any point in time there is only one physically possible future and thus rules out causal indeterminacy, is an empirical claim. However, framing the debate in terms of determinism is a vestige of outmoded Newtonian science. By now the standard interpretation of quantum mechanics holds that determinism is false. But current versions of the so-called determinist views, hard determinism and compatibilism, don't reject quantum indeterminacy. Instead, "determinism" has in effect become a label for the idea that there is no context in which any indeterminacy that does exist is sufficient for control.⁸ Thus the discovery of quantum indeterminacy has not had the impact of refuting the so-called determinist views.

Ironically, it is the naturalized libertarians, more so than the so-called determinists, who are vulnerable to the question of the existence of quantum indeterminacy. For while most who write on free will—including most so-called determinists—accept the standard interpretation of quantum mechanics, this interpretation might be false and the behavior of subatomic particles might be deterministic. If this were so, it would seem to be grounds for rejecting naturalized libertarianism since it would eliminate what seems to be the only indeterminacy purported to be scientifically acceptable.

Nevertheless, this appearance is deceiving. If the standard interpretation of quantum mechanics is false, this would be grounds to reject certain versions of naturalized libertarianism. But it would not affect all versions. For example, it would not affect Timothy O'Connor's version of naturalized libertarianism (discussed below), which doesn't appeal to quantum indeterminacy. Thus, the falsity of the standard interpretation of quantum mechanics would provide only a weak constraint, ruling out only a narrow range of views.

Yet, if the standard interpretation is correct, perhaps empirical considerations can establish a naturalized libertarianism which appeals to quantum indeterminacy.

Kane offers this strategy, attempting to turn the question of libertarian freedom into one for physics and neuroscience.

I will argue, however, that Kane's strategy fails. In fact, resolving the question of the standard interpretation does little to further the free will debate. If the standard interpretation is correct, the question remains whether it makes sense to think that quantum indeterminacy is sometimes sufficient for controlling our actions. If it is false, its falsity provides only a weak constraint.

Kane's Naturalized Libertarianism: Trying to Make Quantum Indeterminacy Matter

Traditionally, libertarianism has rejected the attempt to fit ourselves as free agents into the natural world characterized by science.⁹ Over the past three decades, however, Robert Kane has developed libertarianism in new ways, striving to naturalize it (1996, pp. 17, 115–117; 1999, p. 163). Kane's hope is to lead us out of the mystery of older libertarian views, and he uses the strategy of rendering libertarianism scientifically innocuous as a measure of having done this. However, as Kane recognizes, a naturalized libertarianism continues to face the serious challenge of addressing how libertarian freedom, in holding that indeterminacy is sometimes sufficient for control, makes sense at all. Kane calls the issue of making sense of libertarian freedom the Intelligibility Question (1996, p. 13).

It is crucial to libertarianism that it doesn't hold that just any indeterminacy is sufficient for control. This claim would render libertarian freedom rather blatantly unintelligible. Instead, the libertarian claims that indeterminacy *in a certain context* is sufficient for control. The traditional libertarianism filled out the context (to the extent that this was possible) in terms of a scientifically intractable agent. Naturalizing libertarianism amounts to offering a naturalistic context. And the fundamental problem for naturalized libertarianism is whether it makes sense to think that there is a naturalistic context in which indeterminacy is sufficient for control.

Kane spells out the naturalistic context in terms of locating quantum indeterminacy at a particular point in deliberation involving a conflict of values (such as egoistic and altruistic values). When distinct values support different courses of action in a situation, and a conflict ensues where each course of action is supported by a set of reasons, Kane claims that there is "a kind of stirring up of chaos in the brain that makes it sensitive to the micro-indeterminacies at the neuronal level" (1999, p. 164). His idea is that being torn due to such conflicts creates chaotic conditions which amplify quantum indeterminacy so that its effects percolate up, that is, are manifested at the level of individual neurons, and then at the level of neural networks (1996, pp. 128–130). Kane claims that moral conflicts "create tensions that are

reflected in appropriate regions of the brain by movement further from thermodynamic equilibrium, which increases the sensitivity to micro indeterminacies at the neuronal level and magnifies the indeterminacies throughout the complex macro process which, taken as a whole, is the agent's effort of will" (ibid., p. 130). Thus, Kane's view isn't that moral conflicts create quantum indeterminacies, but rather that they create the conditions in which quantum indeterminacies can be amplified and manifested in our deliberative processes.¹⁰

In particular, this amplification of the effects of quantum indeterminacy makes the outcome of our deliberation indeterminate, since it occurs in the physical basis of the interaction among neurally realized goal-directed states which express our values.¹¹ In addition, the feeling that in such cases branching paths are metaphysically open is a neural sensitivity to these amplified effects of quantum indeterminacy (1996, pp. 130, 132–133; 1999, p. 164). But even though the outcome of deliberation in such cases is indeterminate, it is backed by reasons—since each of the competing courses of action is. Kane's proposal is that the felt indeterminacy and reasons backing of deliberative outcomes is necessary and sufficient for control (1996, pp. 133–135, 141; 1999, pp. 174–176).

The proposal is sketchy and highly speculative. Yet, if Kane successfully addresses the Intelligibility Question, he has converted the issue of libertarian freedom into the empirical question of whether the effects of quantum indeterminacy are manifested in neural processes. If chaotic amplification of the effects of quantum indeterminacy were discovered, then, it seems, libertarianism would receive strong empirical support.

Kane's strategy of naturalizing libertarianism has the effect of focusing our attention on empirical possibilities. Considering the naturalized libertarian's reconciliation of incompatibilism, freedom, and science, this position is attractive. And due to the empirical possibilities and the position's attractiveness, one might be hopeful that Kane's attempt to disarm the Intelligibility Question succeeds.

I will argue, however, that, despite Kane's ingenious strategy, his attempt to disarm the Intelligibility Question and transform the question of libertarianism into one for physics and neuroscience fails. Kane tries to reconcile the arbitrariness of indeterminacy with libertarian freedom as follows:

An ultimate arbitrariness remains in all undetermined SFAs [self-forming actions] because there cannot in principle be sufficient or overriding *prior* reasons for making one set of competing reasons prevail over the other. . . . The absence of an explanation of the difference in choice in terms of prior reasons does not have the tight connection to issues of responsibility one might initially credit it with. . . . None of [the conditions necessary and sufficient for responsibility] is precluded by the absence of an explanation of the difference in choice in terms of prior reasons. (1999, pp. 176–177)

Having our attention distracted by the empirical possibilities, Kane's attempt may seem an adequate answer to the worry about the arbitrariness of indeterminacy which the Intelligibility Question poses. But, at best, Kane shows that indeterminacy is *consistent* with responsibility and, consequently, with the control necessary for responsibility. Thus, he claims, indeterminacy doesn't *preclude* control.¹²

Yet this approach is vulnerable to the following objection. (This objection was raised by Galen Strawson (1986, chapter 2).) Strawson, however, takes as his target libertarianism in general as opposed to just Kane's variety, and Strawson's overall aim is to show that libertarian freedom is incoherent whereas my aim is much more modest—it is merely to show that Kane's argument rests on a prior question with respect to introspective evidence.) The libertarian must render intelligible the idea that there is a context for indeterminacy where it is *sufficient* for control, not just one where it is consistent with control. Kane's approach involves enriching this context through the inclusion of the reasons backing of deliberative outcomes. But the question then becomes whether this inclusion allows us to understand Kane's claim of sufficiency by smuggling in the satisfaction of compatibilist sufficient conditions for control while pointing to an indeterminacy which, irrelevant to sufficiency for control, merely plays the role of satisfying the incompatibilist intuition that indeterminacy is necessary for control.

To make sense of libertarian freedom, and to free libertarianism of the charge that it is (in Galen Strawson's phrase) "a covert compatibilism with an idle incompatibilist premise danglely subjoined" (1986, p. 32), Kane needs to demonstrate that the context for indeterminacy can be enriched in such a way that it is clear that indeterminacy has a role to play in sufficiency for control.

As a result, even if chaotic amplification of indeterminacy were found, the question would remain as to whether we can make sense of naturalized libertarian freedom. Alternatively, as I will discuss below, even if chaotic amplification of quantum indeterminacy were *not* found, the question would remain as to whether a naturalized libertarianism is true. Consequently, physical and neurophysiological research into chaotic amplification of quantum indeterminacy would provide only a weak constraint.

Resolving the question of quantum indeterminacy, either with regard to the correctness of the standard interpretation of quantum mechanics or with regard to chaotic amplification, does little to further the debate. If chaotic amplification of quantum indeterminacy exists, the question remains whether it makes sense to think that there is a context in which quantum indeterminacy is sufficient for control. If it doesn't exist, its absence provides only a weak constraint.

Kane's Appeal to Introspection

However, Kane suggests that enriching the context for indeterminacy must also include the first-person perspective:

... when described from a physical perspective alone, *free will looks like chance*. But the physical description is not the only one to be considered. ... the undetermined outcome of the process, [whether it is settled] one way or the other, is, experientially considered, the agent's choice. (1996, p. 147)

The claim is that when considered from the first-person perspective, it is clear that indeterminacy has a role to play in sufficiency for control. Thus, according to Kane, the contentions of Banks (this volume) and Prinz (this volume) that physical indeterminacy has no such role simply fail to consider this indeterminacy from the first-person perspective of introspection. But for this reasoning to be compelling, we have to decide as a prior issue how heavily we can rely on indetrospective evidence.

Furthermore, as I will show in the next section, Timothy O'Connor's defense of a version of naturalized libertarianism, which provides the major alternative to Kane's, also relies on an appeal to introspection. Thus a broad consideration of the tenability of naturalized libertarianism must take up the question of the accuracy of introspection.

3 O'Connor's Naturalized Libertarianism

O'Connor (2000, p. 67) claims to offer a characterization of libertarian freedom which is consistent with science, joining Kane in attempting to avoid the mystery of a scientifically intractable agent. Thus, he purports to offer a version of naturalized libertarianism. But O'Connor finds Kane's way of spelling out the naturalistic context for indeterminacy—in terms of the physical basis of the interaction among goal-directed states—unconvincing as an attempt to make sense of naturalized libertarian freedom (2000, pp. 36–42). Instead of attempting to explain libertarian freedom in terms of a certain context for quantum indeterminacy, O'Connor's strategy is to tailor his proposal to match our feeling of freedom as exactly as possible, and contend that agents *create* indeterminacy.

O'Connor's view is that free will is a macrodeterminative emergent property of agents, where a macrodeterminative emergent property is "a qualitatively new, macro-level feature" of the world (2000, p. 111). While he claims that such properties are "completely dependent on some set of properties or disjunctive range of properties in the [instantiating] object's microstructure" nevertheless, "[they] exert a causal influence on the micro-level pattern of events that is not reducible to the *immediate* causal potentialities of the subvening properties" (ibid., pp. 111–112). Thus, such properties are "radically new features of the world, in a sense 'transcending' the lower level properties from which [they] emerge" (2000, p. 112). In response to the concern that transcendence indicates that macrodeterminative emergent properties are not naturalistic, O'Connor counters that we need to expand our conception of scientifically acceptable causation to include causation at the

macro-level which is not reducible to the causal properties of micro-level parts (top-down causation) as well as causation at the macro-level which is reducible (bottom-up causation) (2000, pp. 115, 125).

But rather than address the Intelligibility Question, O'Connor's characterization of libertarian freedom as a macrodeterminative emergent property offers a promissory note. Furthermore, there is a serious worry as to whether payment can be made good. While he holds out the prospect that empirical research may discover other examples of macrodeterminative emergence, forcing us to enrich our conception of scientific causation (2000, pp. 110–115),¹³ he provides no reason at all to think that we will be able to understand libertarian freedom as a macrodeterminative emergent property.

O'Connor claims that we are "not wholly moved movers" (2000, p. 67). This, of course, means that we are partly unmoved movers (*ibid.*, p. 97); we are partly moved as well because our reasons influence our choices. Since O'Connor's case for macrodeterminative emergent properties appeals to the possibility of other examples of such properties than libertarian freedom, simply being a macrodeterminative emergent property isn't sufficient for libertarian freedom. Rather, libertarian freedom is a macrodeterminative emergent property of a distinctive sort: a macrodeterminative property that allows unmoved moving. Setting aside the question of whether macrodeterminative emergent properties are consistent with science, O'Connor gives no reason to think that unmoved moving is.

O'Connor's view is important, though, not only because it is a proposal of a purported naturalized libertarianism which doesn't appeal to quantum indeterminacy, but also because, as O'Connor points out, agent causation, in which agents create indeterminacy, is better tailored to the introspective evidence than Kane's view. As O'Connor states, we experience ourselves as creating gaps in causal necessitation by undeterminedly and directly forming a particular intention (2000, p. 124).

For O'Connor, as well as for Kane, the introspective evidence holds out the hope that libertarian freedom is intelligible. This reliance on introspection points to a powerful empirical constraint. I will argue that, because it directly addresses the Intelligibility Question, the question on which the tenability of naturalized libertarianism turns, psychological research into the accuracy of introspection is the source of the most powerful empirical constraint for the problem of free will.

4 Naturalized Libertarianism and Psychological Constraints

Introspection does seem to indicate that when we struggle with a conflict of values, branching paths are metaphysically open. Whatever the metaphysics turns out to be, we experience ourselves as undeterminedly and directly forming the particular intention which causes our action. Thus we feel that we could have done otherwise

(in the same circumstances) if we had formed a different intention. This feeling of freedom is a type of introspective state—that is, it represents a mental state rather than the world—and this type of introspective state provides evidence in favor of libertarian freedom. Kane and O'Connor claim this introspective evidence strongly suggests that, despite our present inability to address the Intelligibility Question, libertarian freedom is nonetheless intelligible.¹⁴

Furthermore, aside from the fact that both Kane and O'Connor appeal to introspective evidence, all libertarians assign introspective evidence some role, for it is our feeling of metaphysically open branching paths that is the *raison d'être* of libertarian freedom. It is the first-person perspective that at least seems to give self-formation meaning. Even if the naturalized libertarian has yet to achieve the goal of making sense of self-formation from the third-person perspective of science, we will not be in a position to confidently conclude that libertarian freedom is *unintelligible* until we address the accuracy of this introspective evidence.

Considering a case of a conflict of values, the libertarian's claim is that it is metaphysically indeterminate which intention is formed. According to the libertarian, some of our intentions are formed in a way that is not merely the result of mental causation, that is, causation by further mental states that are ultimately caused by genetics and environment, but is also the result of indeterminacy. Introspective evidence supports this claim by indicating an absence of sufficient mental causation. The question is whether this introspective evidence is accurate, or whether background beliefs and desires missed by introspection are causally sufficient with respect to which intention produces action. O'Connor suggests that introspection is accurate—that our experience of ourselves as undeterminedly and directly forming a particular intention is as accurate as our perception of ordinary physical objects:

... in the deliberate formation of an intention, the coming to be of my intention doesn't seem to me merely to occur at the conclusion of my deliberation; I seem to experience myself directly bringing it about. . . . This apparent perception of [undetermined] causality could be mistaken, of course; our experience cannot furnish the basis of a 'demonstrative proof' of our having free will. By the same token, our apparent perception of ordinary physical objects also could (in this epistemic sense) be mistaken, yet few take that as a reason not to believe that we do perceive them. (2000, p. 124)

But any answer to the question of the accuracy of introspection is likely to be quite complex, taking into account such factors as the kind of mental state represented (for example, whether it is an intention or a background desire), and empirical research is needed to tease out the complexities. In this effort, psychological research could show that introspection provides a reliable indicator of (the existence or absence of) sufficient mental causation of intention, so that introspective evidence of an absence of sufficient mental causation should be taken seriously. This finding could at least spur us to continue to address the Intelligibility Question.

Or such research could show not only that introspection provides a poor indicator of mental causation of intention but also why it produces illusions of an absence of sufficient mental causation, so that introspective evidence of such an absence should be discarded. In this way, the strongest evidence in support of libertarian freedom would be undermined. As a result, the naturalists among us would have to take seriously the conclusion that libertarianism is refuted.

In addition, this research would provide a payoff with respect to a connection that some, for example Kane, draw between libertarian freedom and consciousness. According to Kane, in cases of libertarian freedom

indeterminism and the effort [of will] are 'fused': the indeterminacy is a property of the effort and the effort is indeterminate. To fully understand how this fusion could take place would be . . . to understand the nature of conscious experience and its unity . . . as well as to understand how consciousness and mind are related, if at all, to the indeterminacy of natural processes. . . . (1998, p. 151)

While Kane rightly points out that understanding quantum indeterminacy and consciousness are problems for everyone, not just libertarians (1998, p. 151), his proposal of a fusion of indeterminacy and conscious effort of will is meant to make sense of libertarian freedom in particular. Given the heady level of speculation, however, the situation cries out for empirical constraints. Kane's proposal of this fusion presents the possibility that indeterminacy and consciousness combine to form something that on the face of it is unintelligible: control for which indeterminacy is necessary and sufficient. If introspection provides a reliable indicator of (the existence or absence of) sufficient mental causation of intention, then this possibility should be taken seriously. But if introspection provides a poor indicator of mental causation of intention, it may be that the introspective evidence in support of libertarian freedom is systematically illusory. If, in addition, research were to explain introspective illusions of an absence of sufficient mental causation, the idea that indeterminacy and consciousness combine to form libertarian freedom would be undermined. While consciousness may play a role in causing behavior (perhaps along the lines carefully and compellingly argued by Pacherie in this volume), it would not in the way that libertarians envision.

Wegner's *Illusion of Conscious Will* amasses psychological research which supports the idea that introspection is a poor indicator of mental causation, at least in the case of causation of action.¹⁵ Wegner's main focus is not the problem of free will but the problem of whether the *feeling* of mental causation of action accurately tracks mental causation of action, a problem which intersects with the problem of free will only at certain points. He finds two broad categories of cases where this tracking goes wrong, which I will call type I and type II inaccuracies (Wegner 2002, pp. 8–11).

Type I inaccuracy (false alarm or false positive) One has a feeling that one's mental states cause action where no such mental causation exists.

Type II inaccuracy (miss or false negative) One does not have a feeling that one's mental states cause action where such mental causation does exist.

Wegner's elaborate "I Spy" experiment is an example of a type I inaccuracy. In the experiment, a subject and an experimental confederate sit at a computer, together controlling a mouse that moves a cursor around a screen which displays a variety of objects (for example, a swan). Both subject and confederate wear headphones; the subject hears 10 seconds of music at the end of each 40-second trial, and, at some point during each trial, also hears a word which in some cases would refer to an object on the screen (for example, "swan"). Subject and confederate are told to stop the cursor sometime during the music. However, the confederate, rather than hearing music or words, hears instructions including, for some trials, instructions on when to stop the cursor. After each stop, the subject rates it according to whether the subject had intended the stop or had allowed it to happen. The experiment found that for trials where the confederate forces the cursor to stop on an object, when the subject hears the word for the object 1 to 5 seconds before the stop (versus 30 seconds before or 1 second after the stop), the subject feels more strongly that he or she had intended the stop (Wegner 2002, pp. 74–78).

This experiment indicates that introspection sometimes misleads us into believing that we have an intention which causes our actions when they aren't caused by our mental states at all (in the experiment, actions are caused by an experimental confederate's forcing the subject to act in certain ways). (However, for reasons to be skeptical about the conclusiveness of this experiment, see Malle's chapter in this volume.) Since the experiment identifies an illusion of control, it might seem to tell against libertarian freedom. But its findings in fact don't help to constrain the problem of free will. For the illusion of control which results from a false alarm isn't relevant to the ordinary example of free will, that is, the case where one's intentions clearly do cause one's actions. Furthermore, the libertarian can accept that there are *some* types of illusions of control; rather, all the libertarian claims is that *not all* types of control are systematically illusory—that is, that there are also some types of control. The relevant question then is whether introspection systematically misses mental causation whenever we have a feeling of libertarian freedom.

Rather, it is the type II inaccuracies, the misses, that are relevant to the question of libertarian freedom, for it is in these sorts of cases that introspection misses mental causation. And, indeed, Wegner shows that these inaccuracies of introspection sometimes mislead us into believing that our mental states are not sufficient causes of our actions when in fact they are. Thus, in these cases, an absence of awareness of sufficient mental causation is erroneously taken to be an awareness of an absence of sufficient mental causation.

These examples include automatism such as alien hand syndrome and automatic writing (Wegner 2002, pp. 4–6, 103–108). In cases of alien hand syndrome, one's hand can do things such as turn pages of a book or unzip a jacket while one isn't aware that one's intentions are sufficient to cause such actions, and one can even think that such intentions have no role in causing such actions. Automatic writing, documented in spiritualist writings in America and Europe in the middle of the nineteenth century, involves writing (including gibberish, reverse writing, or ordinary writing) which one isn't aware of intentionally producing, instead attributing its production to disembodied spirits. These cases vividly show that introspection can miss cases of sufficient mental causation.

Wegner's presentation of research skeptical of the accuracy of introspection is a descendant of Nisbett and Wilson's (1977) influential discussion.¹⁶ Nisbett and Wilson gather research showing that introspection can misidentify or simply miss mental processes, as well as the external inputs and physiological and nonverbal behavioral outputs of mental processes. However, Nisbett and Wilson don't speak directly to the question of whether introspection sometimes misleads us into believing that our mental states don't cause actions when they in fact do. They consider cases where introspection misidentifies mental causes, and they also take up cases where there is an absence of awareness of mental causes (for example, in cases of problem solving), but they don't consider cases where an absence of awareness of sufficient mental causation is mistakenly taken to be an awareness of absence of sufficient mental causation. Wegner's discussion is interesting because it does speak directly to this question.

Yet on the face of it Wegner's examples of automatic writing and alien-hand syndrome have little to do with free will. After all, these are cases where one *lacks* a feeling of control. But the connection with free will is this: type II inaccuracies divide into two subgroups, one group of cases where one feels a lack of control and another where one feels a presence of control. In cases of the first subgroup (type IIa inaccuracies), introspection misses the *intention* producing action. Consequently, the action is not self-attributed and one feels a lack of control. Automatic writing and alien-hand syndrome are examples of type IIa inaccuracies.

However, in cases of the second subgroup (type IIb inaccuracies) introspection misses *mental states such as desires producing the intention* which produces action, rather than the intention itself. (Thus the distinction between subgroups appeals in part to the separation, stressed by both Mele and Malle in this volume, between desires and intentions.) In cases of type IIb inaccuracies, this ignorance of background mental states producing the intention creates the feeling that branching paths are metaphysically open—i.e., that it is metaphysically indeterminate which intention is formed. As a result, we feel that we initiate action in the strong sense of self-formation. Of course, even if introspection does systematically miss back-

ground mental states (an empirical question) and this account of the feeling of libertarian freedom is along the right lines, much more needs to be said, for the feeling withstands the appreciation of this point.

But now it is clear that Wegner's examples are of limited use with respect to the issue of libertarian freedom. Because Wegner's focus is introspective inaccuracies with respect to intentions that cause action, he gives examples of type IIa inaccuracies, but not type IIb inaccuracies, where introspection misses mental causes producing intentions.¹⁷ And even setting this problem aside, Wegner's examples are of the wrong sort. His examples where introspection misses cases of sufficient mental causation involve neuropsychological disorders, as with alien hand syndrome, or specific unusual contexts, such as that set up by the spiritualist movement, which abetted introspective illusions. The libertarian could reasonably contend that the experience of metaphysical indeterminacy in a situation of a conflict of values is neither pathological nor particularly unusual. Wegner's examples might show that introspective evidence is fallible in some unusual cases, but they don't impugn its reliability in more ordinary cases such as the feeling of libertarian freedom.¹⁸

Moreover, the libertarian might claim that if the introspective evidence supporting libertarian freedom were mistaken, this would amount to a systematic illusion. But, the libertarian might assert, such a systematic illusion should be regarded as implausible. Nevertheless, as Wegner notes (2002, p. 137), illusions of the absence of sufficient mental causation can be explained by our background beliefs and desires, as the case of automatic writing in the context of the spiritualist movement suggests. And this explanation might be applied even to systematic illusions. Furthermore, libertarianism's provision of an absolute autonomy suggests a background motivation.¹⁹

Yet even if this point is along the right lines, again, much more needs to be said, for we can appreciate this point and yet the feeling of libertarian freedom remains. In any event, this isn't the place for armchair science.

My conclusion is that the only empirical research which provides constraints for the problem of free will is research relevant to the truth of libertarianism, and, that, because it directly addresses the Intelligibility Question, psychological research regarding the accuracy of introspection offers the most powerful empirical constraint. The libertarian claims that the best explanation of our feeling that there are metaphysically open branching paths is that we become aware of an absence of sufficient mental causes. A specific question for research is whether this is the best explanation. If psychologists were to provide an alternative explanation which not only indicates that there are sufficient mental causes even in ordinary cases where our introspection indicates otherwise, but also offers a model explaining the illusion of their absence, this would undermine any naturalized libertarianism.

Appendix

With respect to other research relevant to the accuracy of introspection, the physiologist Benjamin Libet's research (1985) can be interpreted as indicating that our control of our behavior is limited in a surprising way. Libet's studies (discussed by many of the authors in this volume) indicate the existence of two time gaps (measured in hundreds of milliseconds) in the mental process immediately preceding action: between unconscious brain activity called a readiness potential and the conscious intention to act, and between the conscious intention to act and the muscular motion involved in action. Also, these studies show that during the gap between conscious intention and muscular motion, one can veto the conscious intention and so block action. Furthermore, Libet argues that the veto is itself a "conscious control function" which need not be unconsciously initiated (1999, pp. 558–559).

Libet's finding of the first time gap can be interpreted to give the dizzying impression that our actions are initiated before we consciously intend to act, thus undercutting control at this stage; while the ability to veto conscious intentions during the second time gap provides us with control at a later stage. Thus our control is limited. Yet Dennett (2003, pp. 227–242) argues that this interpretation of the first time gap mistakenly assumes that conscious intentions immediately prior to action exhaust the mental processes involved in control, and thus that such processes don't also include unconscious processes or conscious deliberations long before action.

In any event, it is not clear that the possibility of a surprising limitation of control is a constraint on the problem of free will at all, in that it doesn't seem to rule out any positions on the problem. While Libet's findings show that some of the causes of our actions are not transparent to introspection, the libertarian needn't have claimed otherwise. Furthermore, the power to veto conscious intentions could be co-opted by libertarians as Campbellian resistance of temptation (a vivid image repeated by Kane (1996, pp. 126–128)); for this approach to Libet's data, see Kane 1996, p. 232, note 12. Libet's findings don't rule out any versions of compatibilism or hard determinism either, these views not being committed to the accuracy of introspection. (For extended discussions of reasons to be skeptical about the importance of Libet's findings to the issue of free will, see the chapters in this volume by Gallagher and Mele.)

In fact, Libet's conclusion about empirical constraints on the problem of free will is similar to mine, leaving the introspective evidence of libertarian freedom open to empirical study. After acknowledging that his findings don't rule out free will (1999, pp. 551 and 561), Libet states:

... we must recognize that the almost universal experience that we can act with a free, independent choice provides a kind of prima facie evidence that conscious mental processes can causatively control some brain processes. This creates, for an experimental scientist, more dif-

ficulty for a determinist than for a non-determinist option. The phenomenal fact is that most of us feel that we do have free will, at least for some of our actions and within certain limits that may be imposed by our brain's status and by our environment. The intuitive feelings about the phenomenon of free will form a fundamental basis for views of our human nature, and great care should be taken not to believe allegedly scientific conclusions about them that actually depend upon hidden ad hoc assumptions. A theory that simply interprets the phenomenon of free will as illusory and denies the validity of this phenomenal fact is less attractive than a theory that accepts or accommodates the phenomenal fact. (*ibid.*, p. 563)

Wegner's claim is that the feeling of conscious control is illusory, and his goal is to accommodate—that is, explain in naturalistic terms—the phenomenal fact (2003, chapter 9, especially pp. 318 and 325–334), but, as I have argued, he doesn't provide adequate evidence to establish that conscious control in fact is illusory in a relatively ordinary case such as the feeling of libertarian freedom.

Acknowledgments

My greatest debt with respect to thinking through the issues of this chapter is to Dale Turner, with whom I taught a seminar on free will in the spring of 2003. I presented versions of this chapter at the 2004 APA Pacific Division Meeting in Pasadena and the 2004 Joint Meeting of the Society for Philosophy and Psychology and the European Society for Philosophy and Psychology in Barcelona. I am thankful to audience members at those events for comments, and in particular to Stephan F. Johnson for his comments at the Pasadena meeting. I also owe thanks to Dion Scott-Kakures, Michael Cholbi, and William P. Banks for extremely useful comments on earlier drafts.

Notes

1. See Hardin 1988. Also see Ross 2001 and peer commentary for further discussion.
2. The problem of free will is commonly stated in terms of whether we control our actions. Stated more fully, the problem is whether we control our actions through controlling the goal-directed states—intentions, desires, or other goal-directed states—which produce action. (Gallagher, in this volume, also emphasizes that the problem of free will should be characterized in terms of control of environmentally situated intentional actions rather than control of bodily movement.) If we call the mental capacity for producing goal-directed states "the will," then the problem is whether we control our goal-directed states, and consequently our actions, through controlling the will. (There are other uses of the word 'will'. For example, it is sometimes used to refer to the goal-directed states that are products of the mental capacity called the will—see Frankfurt 1971 for this usage.)
3. According to the claim that our actions are brought about by genetics and environment, even if we can in some ways manipulate our genes through gene therapy and in some ways manipulate our adult environments, ultimately we control neither our genes nor environments. For our manipulation of our genes and adult environments is brought about by goal-directed states, and such states are in turn ultimately brought about by aspects of our genetics and environments which we don't control.
4. For this idea of self-formation, see Kane 1996, pp. 74–77 and 124–125; Kane 1999, pp. 164–165. I have described self formation as a negative claim along with indeterminacy. Kane seems to suggest more of

a positive aspect: "Free will . . . is the power of agents to be the ultimate creators (or originators) and sustainers of their own ends or purposes." (1996, p. 4) But this description isn't helpful until we understand what it means to be an ultimate creator, and Kane's suggestion is largely negative: "Such a notion of *ultimate* creation of purposes is obscure. . . . Its meaning can be captured initially by an image: when we trace the causal or explanatory chains of action back to their sources in the purposes of free agents, these causal chains must come to an end or terminate in the willings (choices, decisions, or efforts) of the agents. . . . If these willings were in turn caused by something else, so that the explanatory chains could be traced back further to heredity or environment, to God, or fate, then the ultimacy would not lie with the agents but with something else." (1996, p. 4) However, Kane is clear that ultimate creation involves metaphysical indeterminacy. Thus, I take the involvement of indeterminacy in the production of actions to be the sole positive aspect of self formation. I will discuss the relation between indeterminacy and control below.

5. Derk Pereboom (2001, pp. 110–117) offers this argument. For a similar argument, see Taylor 1992, pp. 45–47.

6. For example, hard determinism can accept justification for punishment on the basis of social utility. For a discussion of a sophisticated hard determinist's account of ersatz normativity (that is, normativity without self formation), see chapters 5–7 of Pereboom 2001. Prinz (this volume) also comments on the social utility of ersatz ("invented") control.

7. Dennett admits that a sophisticated hard determinist's account of ersatz normativity is only "terminologically different" from compatibilism (2003, pp. 97–98).

8. Hard determinists and compatibilists divide as to whether indeterminacy is necessary for control, the hard determinists claiming that it is and the compatibilists claiming that it isn't. However, both sorts of determinism reject the libertarian's claim that indeterminacy is sometimes sufficient for control.

9. An example of a traditional non-naturalist libertarian view is Roderick Chisholm's proposal that "we have a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved" (1964, p. 34).

10. For a helpful discussion of quantum indeterminacy and chaotic amplification, see Bishop 2002, especially section 3.

11. As far as this discussion is concerned, the goal-directed states that express our values can be characterized as desires or, alternatively, as besires, which are hybrids of beliefs and desires. For discussions of besires, see van Roojen 1995 and Ross 2002.

12. Similarly, Kane states that "the core meaning of 'he got lucky,' which is implied by indeterminism, I suggest, is that 'he succeeded despite the probability or chance of failure'; and this core meaning does not imply lack of responsibility, if he succeeds" (1999, p. 171).

13. Pereboom (2001, pp. 74, 85–86) also considers the existence of top-down causation to be an empirical question, although he is pessimistic about its prospects.

14. Campbell (1957, pp. 176–178) states this point in classic fashion. Also see O'Connor 2000, p. 124. Kane puts the point in a way reminiscent of Campbell: ". . . when described from a physical perspective alone, *free will looks like chance*. But the physical description is not the only one to be considered. The indeterministic chaotic process is also, experientially considered, the agent's effort of will; and the undetermined outcome of the process, [whether it is settled] one way or the other, is, experientially considered, the agent's choice. From a free willist point of view, this experiential or phenomenological perspective is also important; it cannot simply be dispensed with." (1996, p. 147) At points Kane conflates epistemic and metaphysical indeterminacy. Kane states: "Every free choice (which is a [self formed willing]) is the initiation of a 'value experiment' whose justification lies in the future and is not fully explained by the past. It says, in effect 'Let's try this. It is not required by my past, but it is consistent with my past and is one branching pathway my life could now meaningfully take. I am willing to take responsibility for it one way or the other. . . .'" (1996, p. 145) Kane repeats this statement in a later work and continues as follows: "To initiate and take responsibility for such value experiments whose justification lies in the future, is to 'take chances' without prior guarantees of success. Genuine self-formation requires this sort of risk-taking and indeterminism is a part of it. If there are persons who need to be certain in advance just exactly what is the best or right thing to do in every circumstance (perhaps to be told so by some human or divine authority), then free will is not for them." (1999, p. 176) Claiming that my action is "not required by my past" suggests metaphysical indeterminacy. But a lack of certainty with

respect to the best future action is an epistemic matter. This point is important because while everyone has to accept that epistemic indeterminacy is involved in processes of deliberation leading to the production of action—uncertainty about what action is best is a fact of life—it is only the libertarian who claims that metaphysical indeterminacy involved in such processes is sufficient for controlling action. Thus, Kane's conflation tends to give libertarianism more credibility than it deserves. However, Kane's misleading description of the indeterminacy doesn't detract from the point that libertarians use introspective evidence to support metaphysical indeterminacy.

15. In addition to supporting the idea that introspection is fallible, Wegner puts forth the further claim that such introspective states are causally inefficacious (2002, pp. 317–318). For a careful elucidation of Wegner's claim and a convincing argument that it is false, see Nahmias 2002; also see, in the present volume, Bayne's trenchant criticism of this aspect of Wegner's work. Some of Wegner's claims seem to place him in the compatibilist camp (see, for example, pp. 318–319). But because he never makes clear what he means by "control" in the way that philosophers in the free literature do, it is not clear whether Wegner is a compatibilist or a hard determinist.

16. Wegner (2002, p. 67) notes this heritage.

17. In addition, while Wegner (2002, pp. 65–95) offers a psychological model for the experience of mental causation of action, unfortunately it has features that make it specific to causation of action and so doesn't straightforwardly apply to the existence or absence of mental causation of mental states.

18. Wegner (2002, p. 327) acknowledges that introspection is accurate much of the time. While Nisbett and Wilson (1977) offer a more sweeping skepticism about the accuracy of introspection (contending that the accuracy of first-person verbal reports of mental processes is no better than third-person guesses (pp. 248–251)), White's 1988 review of the psychological literature shows that their claims needed to be modified, and that introspection may be highly accurate in some cases (pp. 36–37).

19. Kane notes that his conception of self formation is closely connected to our "life-hopes"—including dispositions to view ourselves as ultimately responsible for our own characters and achievements rather than as pawns of nature, fate, or the whims of others, including God" (1996, p. 4).

References

- Bishop, R. C. 2002. Chaos, indeterminism, and free will. In *The Oxford Handbook of Free Will*, ed. R. Kane. Oxford University Press.
- Campbell, C. A. 1957. Has the self 'free will'? In Campbell, *On Selfhood and Godhood*. Allen and Unwin.
- Chisholm, Roderick M. 1964. Human freedom and the self. The Lindley Lecture. Reprinted in Watson 2003. Page numbers refer to reprint.
- Dennett, Daniel C. 2003. *Freedom Evolves*. Viking.
- Ekstrom, Laura, ed. 2001. *Agency and Responsibility: Essays on the Metaphysics of Freedom*. Westview.
- Frankfurt, Harry 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68: 5–20.
- Hardin, C. L. 1988. *Color for Philosophers: Unweaving the Rainbow*. Hackett.
- Kane, R. 1996. *The Significance of Free Will*. Oxford University Press.
- Kane, R. 1999. Responsibility, luck, and chance: Reflections on free will and indeterminism. *Journal of Philosophy* 96: 217–240. Reprinted in Ekstrom 2001. Page numbers refer to reprint.
- Kane, R., ed. 2002. *The Oxford Handbook of Free Will*. Oxford University Press.
- Libet, B. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* 8: 529–566.
- Libet, B. 1999. Do we have free will? *Journal of Consciousness Studies* 6, no. 8–9: 47–57. Reprinted in Kane 2002. Page numbers refer to reprint.
- Nahmias, E. 2002. When consciousness matters: A critical review of Daniel Wegner's *The Illusion of Conscious Will*. *Philosophical Psychology* 15: 527–541.
- Nisbett, R. E., and Wilson, T. D. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84: 231–259.

- O'Connor, T. 2000. *Persons and Causes: The Metaphysics of Free Will*. Oxford University Press.
- Pereboom, D. 2001. *Living without Free Will*. Cambridge University Press.
- Ross, P. W. 2001. The location problem for color subjectivism. *Consciousness and Cognition* 10: 42–58.
- Ross, P. W. 2002. Explaining motivated desires. *Topoi* 21, no. 1–2: 199–207.
- Strawson, G. 1986. *Freedom and Belief*. Oxford University Press.
- Taylor, R. 1992. *Metaphysics*, fourth edition. Prentice-Hall.
- van Roojen, M. 1995. Humean motivation and Humean rationality. *Philosophical Studies* 79: 37–57.
- Watson, G. ed. 2003. *Free Will*, second edition. Oxford University Press.
- Wegner, D. M. 2002. *The Illusion of Conscious Will*. MIT Press.
- White, P. A. 1988. Knowing more about what we can tell: "Introspective access" and causal report accuracy 10 years later. *British Journal of Psychology* 79: 13–45.

8 Toward a Dynamic Theory of Intentions

Elisabeth Pacherie

In this chapter, I shall offer a sketch of a dynamic theory of intentions. I shall argue that several categories or forms of intentions should be distinguished based on their different (and complementary) functional roles and on the different contents or types of contents they involve. I shall further argue that an adequate account of the distinctive nature of actions and of their various grades of intentionality depends on a large part on a proper understanding of the dynamic transitions among these different forms of intentions. I also hope to show that one further benefit of this approach is to open the way for a more perspicuous account of the phenomenology of action and of the role of conscious thought in the production of action.

I take as my point of departure the causal theory of action (CTA). CTA is the view that behavior qualifies as action just in case it has a certain sort of psychological cause or involves a certain sort of psychological causal process. In the last decades, CTA has gained wide currency. Yet it covers a variety of theories with importantly different conceptions of what constitutes the requisite type of cause or causal process qualifying a piece of behavior as an action. Broadly speaking, CTA takes actions to be associated with sequences of causally related events and attempts to characterize them in terms of certain causal characteristics they have. Versions of CTA can take different forms depending on what they take the elements of the action-relevant causal sequence to be and on what part of the sequence they identify as the action.

The earlier belief/desire versions of CTA, made popular most notably by Davidson (1980, essay 1) and Goldman (1970), held that what distinguishes an action from a mere happening is the nature of its causal antecedent, conceived as a complex of some of the agent's beliefs and desires. However, it soon appeared that simple belief/desire versions of the causal theory are both too narrow and too unconstrained. On the one hand, they do not deal with "minimal" actions, those that are performed routinely, automatically, impulsively, or unthinkingly. On the other hand, they are unable to exclude aberrant manners of causation when specifying the