Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
Vol. LXXXVIII No. 2, March 2014
doi: 10.1111/j.1933-1592.2012.00609.x

Explaining Away Incompatibilist Intuitions

DYLAN MURRAY
University of California, Berkeley

EDDY NAHMIAS*
Georgia State University

The debate between compatibilists and incompatibilists depends in large part on what ordinary people mean by 'free will', a matter on which previous experimental philosophy studies have yielded conflicting results. In Nahmias, Morris, Nadelhoffer, and Turner (2005, 2006), most participants judged that agents in deterministic scenarios could have free will and be morally responsible. Nichols and Knobe (2007), though, suggest that these apparent compatibilist responses are performance errors produced by using concrete scenarios, and that their abstract scenarios reveal the folk theory of free will for what it actually is—incompatibilist. Here, we argue that the results of two new studies suggest just the opposite. Most participants only give apparent incompatibilist judgments when they mistakenly interpret determinism to imply that agents' mental states are *bypassed* in the causal chains that lead to their behavior. Determinism does not entail bypassing, so these responses do not reflect genuine incompatibilist intuitions. When participants understand what determinism does mean, the vast majority take it to be compatible with free will. Further results indicate that most people's concepts of choice and the ability to do otherwise do not commit them to incompatibilism, either, putting pressure on incompatibilist arguments that rely on transfer principles, such as the Consequence Argument. We discuss the implications of these findings for philosophical debates about free will, and suggest that incompatibilism appears to be either false, or else a thesis about something other than what most people mean by 'free will'.

1. Investigating the Meaning of 'Free Will'

Debates about free will remain mired in "dialectical stalemates" (Fischer 1994). Compatibilists typically agree that *if* free will were what incompatibilists say it is—a type of freedom that requires having an

* Authorship is equal.

unconditional ability to do otherwise or being the "ultimate source" of one's actions—then it would be incompatible with determinism. Incompatibilists typically agree that *if* free will were what compatibilists say it is—a type of freedom that requires a less metaphysically demanding set of capacities, such as reflective, rational self-regulation of one's actions—then it would be compatible with determinism. Each side believes that the other is wrong about what free will is, and about what conditions are required for having it, but they agree on which conditions are compatible with determinism and on which are not. To avoid this morass, we might try banning the term 'free will' from discussion and proceeding instead by using the terms 'freedom$_I$' and 'freedom$_C$' for incompatibilists' and compatibilists' respective conceptions of it (Chalmers 2011). Doing so would likely avoid some confusion. But the debate would surely persist, because much of it *is* verbal—not in any pejorative sense, but in that much of the fundamental impasse between compatibilists and incompatibilists just is over what we *mean* by 'free will'; about which conception of free will our inquiry concerns.

We take it that philosophical investigations of concepts used in everyday, nonphilosophical life are typically concerned with, and are at least importantly constrained by, the ordinary or "folk" understanding of those concepts. This is certainly the case when the concepts are normative. The default method for theorizing about many normative concepts, wide reflective equilibrium (WRE), takes as inputs our normative principles, background scientific theories, and pre-theoretical (but reflective) judgments, or intuitions, about relevant cases, and then attempts to develop a philosophical theory that is maximally consistent (and, ideally, mutually justifying) among those inputs.[1] WRE builds upon the pre-theoretical understanding of the concept under investigation, as revealed by intuitions about specific cases, so that our final theory—even if it deviates from that understanding to reach equilibrium with other inputs—is recognizable and relevant to the normative roles that the concept plays in everyday human life.

'Free will' plays a central role in the conceptual scheme that we use to navigate the normative world via its connections to 'moral responsibility', 'blame', 'autonomy' and related concepts. Theorizing about 'free will' in isolation from the ordinary understanding of it thus risks being an academic exercise about some other, technical conception with understanding of it divorced from people's actual practices of assessing praise, blame, reward, and punishment, and from their understanding of them-

[1] See Goodman (1955), Rawls (1971), and Daniels (1979). For application of WRE to debates about free will and moral responsibility, see Fischer and Ravizza (1998, 10–11), Swanton (1992), and Vargas (2010).

selves and their place in the world. Imagine, for instance, that freedom_C (and not freedom_I) is what ordinary people care about having, and worry about not having—that freedom_C is what they refer to when thinking about, speaking of, and otherwise using the concept 'free will'. If so, would the fact that freedom_I is incompatible with determinism support any interesting or important version of incompatibilism? We cannot see how. Philosophers are of course free to discuss some other, stipulated notion of 'free will', but for philosophers interested in escaping the dialectical stalemates, the starting point that compatibilists and incompatibilists should both be able to agree on is figuring out what people mean by 'free will'—freedom_I or freedom_C.[2]

We can investigate the ordinary understanding of free will in several ways. First, we can make "field observations" of how people employ the concept in everyday interactions—under what circumstances they claim that one has, or has not, acted of one's own free will. Peter Strawson (1962, 87) appeals to this method in observing that when we actually suspend a reactive attitude, such as praise or blame, "it is *never* the consequence of the belief that the piece of behaviour in question was determined." When people speak of free will and moral responsibility in the course of everyday life, he claims, the contrast class is with actions produced by mental incapacity, ignorance, coercion, and the like, never with determinism.[3]

---

[2] R. Jay Wallace (1994), drawing on Strawson (1962), argues that the central debate between compatibilists and incompatibilists is a normative dispute about what understanding of free will would make it fair to hold people morally responsible—a question which cannot be settled by facts external to, and independent from, our moral practices. There may even be no such independent facts. On this interpretation, incompatibilists claim that our moral practices are committed to its being unfair to hold people morally responsible if determinism is true. See also Vargas (2004). We share this general approach to the debate; if we try to figure out what free will is without first figuring out which properties are relevant to our normative practices, we may do interesting metaphysics, but not necessarily the metaphysics of *free will*. Similar points apply to recent scientific attempts to show that free will is an illusion (e.g., Wegner 2002, Bargh 2008, and Harris 2012). These authors only reach their conclusion by stipulating a specific definition of free will, which they typically simply presume is the ordinary understanding. These authors would be better advised to begin by actually investigating that assumption. If their definitions are not accurate representations of the ordinary understanding of free will, then certain discoveries might show that something is an illusion, but not that free will of the sort under discussion is (see Nahmias, forthcoming). Finally, we should emphasize that our results are not only relevant to philosophical and scientific debates on the approach to the problem of free will advocated above. WRE and other approaches that begin by investigating the roles it plays provide familiar frameworks in which the need for information about the ordinary concept of free will is most apparent, but there are many other reasons to pursue such an investigation, as well (see, e.g., Nahmias et al., 2006).

[3] On the other hand, Galen Strawson claims, "it is also in our nature to take determinism to pose a serious problem for our notions of responsibility and freedom" (1986, 89).

Second, we can simply ask people what they mean by 'free will'. Monroe and Malle (2009) employ this method by giving participants a free response task in which they are asked to define what free will is. Most participants' responses mentioned the ability to make choices, to act on one's desires, and to be free from specific external constraints, whereas almost none appealed to any considerations regarding determinism or indeterminism.

While these approaches offer some insight, they have their limitations, and we should not conclude on their basis alone that the ordinary concept of free will refers to freedom_C. People's concept of free will might implicitly involve some relation to determinism that people are not explicitly aware of, such that they do not mention determinism unprompted. As Frank Jackson (1998) argues, philosophical thought experiments are designed to tap into just these sorts of tacit presuppositions or intuitions. This "method of possible cases" elicits information about our concepts that observations of ordinary linguistic usage typically do not. The Gettier cases, for instance, did not convince us to reform our concept 'knowledge'. Rather, they showed "that it had never been true justified belief that was the crucial factor, but it took the cases to make this obvious, to make explicit what had been implicit in our classificatory practice all along" (Jackson 1998, 36).[4] In the same way, intuitions evinced by the method of possible cases can bring to light the tacit commitments of the ordinary concept of free will.

The method of possible cases needs to be employed with some care and sophistication, however, in order to ensure that it elicits intuitions about the concept actually being investigated. As Jackson (1998, 35) warns, when investigating people's concept '*K*':

> A person's first-up response as to whether something counts as a *K* may well need to be discounted. One or more of: the theoretical role they give *K*-hood, evidence concerning other cases they count as instances of *K*, signs of confused thinking on their part, cases where classification is, on examination, a derivative one (they say it's a *K* because it is very obviously a *J*, and they think, defeasibly, that any *J* is a *K*), their readiness to back off under questioning, and the like, can justify rejecting a subject's first-up classifications as revealing their concept of *K*-hood.

We believe that this is useful advice and that experimental philosophy, by presenting surveys of possible cases to non-philosophers, is well-situated to act on it.

[4] Though see Weinberg et al. (2001) for experimental results that may cast doubt on how widespread the Gettier intuitions are. See also note 5.

Employing the method of possible cases experimentally allows comparison of how people's responses differ depending on subtle, controlled differences between cases, thereby offering insights into the psychological processes that generate people's intuitions and into which particular features of cases those intuitions "track" or are influenced by (Sripada and Konrath 2011; Sripada forthcoming). It also allows researchers to test how people will "back off under questioning" in order to achieve reflective equilibrium between seemingly inconsistent sets of responses (Nichols and Knobe 2007; Lombrozo 2009), and to explore the theoretical roles that people's concepts play (e.g., their connection to other concepts). For example, people might not initially recognize that free will requires an unconditional ability to do otherwise (i.e., holding fixed all prior conditions and laws), but they might think that the ability to make genuine choices does require such an ability, and genuine choices might be shown to be central to their concept of free will. In that case, we could discover that people's concept of free will involves incompatibilist commitments even though they do not initially recognize as much (see section 4 below; see Sommers 2010).

Experimental philosophy can, in these ways, correct assumptions about what is intuitive to non-philosophers and clarify when, and to what extent, philosophical theories are *systematizing* beliefs, concepts, and intuitions, or conversely, when they are suggesting their *revision* (see Vargas 2009). Because we need ordinary understandings of concepts to constrain WRE, and because intuitions can serve as guides to what they are, we should supplement philosophical analysis by employing the method of possible cases in the most systematic fashion available, using participants who have no direct stake in the philosophical debates.[5]

Traditionally, incompatibilists have claimed that their position is more intuitive to non-philosophers than compatibilists'. Peter van Inwagen (2009, 257), for example, claims that:

> It has seemed obvious to most people who have not been exposed (perhaps 'subjected' would be a better word) to philosophy that free

[5] Some take experimental philosophy to undermine traditional philosophical methodology, such as conceptual analysis and reflective equilibrium, by casting doubt on whether context-invariant, "real intuitions" exist at all (see, e.g., Stich 1988, Stich and Weinberg 2001, Weinberg et al. 2001, and Weinberg 2007). Whether there are "real intuitions" about any given subject matter, though, is an empirical question that must be examined on a case-by-case basis. Studies have shown considerable variability in some intuitions, but taking the extant evidence to suggest that all intuitions are problematically variable is inductively unwarranted. The data we present below provide some evidence that there are intuitions that can serve as reliable inputs to WRE regarding free will.

> will and determinism are incompatible. It is almost impossible to get beginning students of philosophy to take seriously the idea that there could be such a thing as free will in a deterministic universe. Indeed, people who have not been exposed to philosophy usually understand the word 'determinism' (if they know the word at all) to stand for the thesis that there is no free will. And you might think that the incompatibility of free will and determinism deserves to be obvious—because it *is* obvious.[6]

Van Inwagen raises another way in which the method of possible cases can go astray here, one which experimental philosophy can test and control for. If people take free will to be incompatible with "the thesis that there is no free will," that would not support incompatibilism. 'Determinism', as employed in the philosophical debate about free will, is the thesis that, necessarily, a complete description of the state of the universe at one time and of the laws of nature logically entails a complete description of the state of the universe at any other time (van Inwagen 1983). Compatibilists and incompatibilists agree that this (or some similar version of determinism) is the thesis at issue in the philosophical debate, and that the intuitive incompatibility of free will and other theses labeled "determinism" provides no support for incompatibilists.[7] That determinism might be confused with other theses, however, might make it look as though people have incompatibilist intuitions when they in fact do not.

Indeed, compatibilists have a long history of accusing incompatibilists of conflating determinism with other theses. Chrysippus charged proponents of the "lazy argument" of confusing determinism with fatalism, noting that it does not follow from its being determined whether or not one will die, for example, that one will die whether or not one goes to the doctor (O'Keefe 2005, 145–147). Schlick (1939) warned against conflating the laws of nature with prescriptive laws that have genuine coercive force—the former do not compel one's compliance in the way that the laws of an oppressive government might.

---

[6] Similar claims are made by Kane (1999, 217, also citing William James and Kant; 2005, 12–13), O'Connor (2000, 4), Strawson (1986, 89), Ekstrom (2002, 310), Pereboom (2001, xvi), Pink (2004, 12), and Cover & Hawthorne (1996, 51).

[7] Of course, it may be that the intuitive incompatibility of free will and some other thesis labeled "determinism," such as theological, logical, or psychological determinism, is what originally gave rise to the philosophical debates. Nonetheless, the contemporary debate concerns determinism of the sort defined above. As such, intuitions concerning the incompatibility of any other form of "determinism" and free will do not provide any direct evidential support for one side over the other in the current philosophical debate.

Another thesis people might conflate determinism with is *bypassing,* which occurs when one's actions are not causally dependent on one's relevant mental states and processes, such as one's beliefs, desires, deliberations, and decisions.[8] An agent's mental states are bypassed when she ends up doing what she does regardless of what they were. One form of bypassing is *fatalism*, the thesis that certain things will happen *no matter what* one wants, decides, or tries to do, such that nothing could happen other than what actually happens *even if* one's past mental states, such as one's desires, had been different. Another form of bypassing is *epiphenomenalism*, the thesis that one's (conscious) mental events or processes have no causal effect on physical events, including one's behaviors.

Bypassing can also come in more specific forms, such as the causal irrelevance of psychological capacities specifically highlighted by compatibilists—e.g., capacities to be responsive to reasons (Fischer 1994), or specifically moral reasons (Wallace 1994), capacities to act in accord with higher-order volitions (Frankfurt 1971), or capacities for reflective self-governance (Scanlon 1998). But if an agent's general psychological states and processes are bypassed, then these more specific compatibilist capacities will be bypassed, as well. Lack of bypassing is necessary for free and responsible agency according to nearly every philosophical account of free will, compatibilist and incompatibilist, which all agree that free actions must be causally dependent on one's decisions, desires, beliefs, and deliberations.

Compatibilists and incompatibilists also agree that determinism, properly understood, does *not* involve or entail bypassing. One's mental states are not excluded from the causal chains that lead to one's actions just because the past and the laws of nature are sufficient for those actions' occurrence. The past and the laws of nature may completely cause one's actions *via* causing intermediary mental states, in which case those mental states are the proximal causes of one's actions and are *not* bypassed. (Nor, of course, does bypassing entail determinism.) That determinism and bypassing are distinct, though, may not be clear to non-philosophers.

---

[8] Related use of the term "bypassing" was introduced in Blumenfeld (1988) and Mele (1995). Precisely what kind of causal dependence and what the relevant mental states are that bypassing precludes will have to be provided by a final theory of intentional mental causation. A full definition of bypassing would also need to incorporate a clause to the effect that the dependence is non-deviant, and should also allow that bypassing comes in degrees rather than being all-or-nothing. We set aside these complications here. While the lack of bypassing is necessary for free will, we do not suggest that it is sufficient for it.

Van Inwagen and many others claim that incompatibilism is intuitive, and the results of some experimental philosophy studies (e.g., Nichols and Knobe 2007) have seemed to support that contention. But if people often conflate determinism with bypassing, then what seem to be intuitions that free will is incompatible with determinism may instead be intuitions that free will is incompatible with bypassing. As we explain in the next section, there is reason to worry that the results of previous studies have been contaminated by exactly that mistake. In sections 3 and 4, we attempt to follow Jackson's advice—to use the method of possible cases to test and control for that confound. In section 5, we discuss the implications for the philosophical debate.

2. Previous Experimental Philosophy on Free Will

Previous research on whether most people have incompatibilist intuitions has yielded conflicting results (see, e.g., Nichols 2011). In studies by Nahmias, Morris, Nadelhoffer, and Turner (NMNT) (2005, 2006), 65–85% of participants judged that agents in deterministic scenarios could act of their own free will, be morally responsible, and deserve praise and blame for their actions. Nichols and Knobe (N&K) (2007), however, suggest that these results may be biased by using concrete scenarios that trigger strong emotions. N&K present evidence that 86% of participants judged that it is not possible for an agent to be "fully morally responsible" in an abstract deterministic scenario that does not specify a particular agent or action. In contrast, in their concrete deterministic scenario, 72% of participants said that a specific agent, Bill, was "fully morally responsible" for killing his family in order to have an affair with his secretary. N&K suggest that the best explanation of both sets of results—theirs and NMNT's—is that concrete cases engage people's emotions in a way that abstract cases do not, thereby leading participants to offer *apparent* (or in Jackson's terms, "first-up") compatibilist judgments. According to this *affective performance error model*, these compatibilist responses are "performance errors brought about by affective reactions. In the abstract condition, people's underlying theory is revealed for what it is—incompatibilist" (2007, 672).[9]

[9] N&K are tentative about the model and discuss several others, including an affective competence model, concrete competence model, and hybrid models, but they claim that "the experimental evidence gathered thus far seems to suggest that the basic idea behind [the affective performance error] model is actually true" (2007, 678).

However, the affective performance error model simply cannot account for all of the previous results. The majority of participants in NMNT's (2005, 2006) studies gave compatibilist responses even to scenarios that *did not involve high affect*—some of which involved positive actions (saving a child from a burning building; returning money one finds in a lost wallet) and some of which involved affectively neutral actions (going jogging). No significant difference in responses about free will and moral responsibility was found between these scenarios and those that involved negative actions (robbing a bank, stealing a necklace, and keeping the money found in a lost wallet).[10] Moreover, NMNT's (2006) scenarios that did involve negative actions were not especially high-affect. Stealing a necklace, robbing a bank, and keeping $1000 found in a wallet seem emotionally more similar to N&K's (2007) low-affect condition, in which the agent cheats on his taxes, than to N&K's high-affect condition, in which Bill burns his family to death so that he can have an affair with his secretary.

That most participants in NMNT's studies (2005, 2006) gave compatibilist responses to low-affect (and affect-neutral) scenarios speaks against the performance error model, since it shows that it is not the case that most people give compatibilist responses only when a scenario involves high negative affect. Some other explanation for the difference in responses in N&K's and NMNT's studies is required. We agree that high affect can bias participants' responses, and suspect that it does so in N&K's very high-affect concrete case, but we do not believe that this bias is responsible for compatibilist judgments in general. In fact, we suspect that most of the *incompatibilist* judgments elicited in N&K's study are the product of a different sort of error—namely, participants' interpreting the description of determinism to involve bypassing, leading them to give apparent incompatibilist responses when in fact their underlying concept is not incompatibilist.

N&K's description of determinism (see below) states that in Universe A, "given the past, each decision *has to happen* the way that it does," and it ends by contrasting this universe with Universe B, in which "each human decision *does* ***not*** *have to happen* the way that it does." Participants may interpret this wording to imply that agents' decisions are bypassed in Universe A if they take it to mean that each decision has to happen *no matter what*—that is, that each decision has to happen the way it does regardless of what happened in the past,

[10] In the "Jeremy" scenario, affirmative responses for free will were 76% (negative action), 68% (positive), and 79% (neutral), and for moral responsibility 83% (negative) and 88% (positive). In the "Fred and Barney" scenario, affirmative responses for free will were 76% (negative action) and 76% (positive), and for moral responsibility 60% (negative) and 64% (positive) (see Nahmias et al. 2006, 39).

including what beliefs and desires the agent had leading up to the decision.[11]

N&K claim that "one cannot plausibly dismiss the high rate of incompatibilist responses in the abstract condition as a product of some subtle bias in our description of determinism" because "the concrete condition used precisely the same description, and yet subjects in that condition were significantly more likely to give compatibilist responses" (2007, 670–71). However, their description of determinism might have misleading features that *interact* with other features unique to either the abstract or concrete case. If the high negative affect in N&K's concrete case biases people to judge that Bill is morally responsible for killing his family, it may also lead many participants to neglect features of the scenario that might otherwise mitigate their responsibility attributions, such as bypassing. In other words, N&K's description of determinism may lead people to interpret it to involve bypassing (in both the abstract and concrete cases), but this mistake may be "cancelled out" in the concrete, but not in the abstract case, because of the latter's high negative affect (Bill's killing his family).

Nahmias, Coates, and Kvaran (2007) provided some initial evidence for this "debunking explanation" of people's apparent incompatibilist intuitions. In their studies, most participants responded that agents in a deterministic universe could have free will, be morally responsible, and deserve blame *if* the scenario described the agents' decisions as being "completely caused by the specific thoughts, desires, and plans occurring in their minds." When agents' decisions were described as being "completely caused by the specific chemical reactions and neural processes occurring in their brains," on the other hand, most people judged that the agents could *not* be free or responsible.[12] The latter, reductionistic description may lead people to take agents' mental states

---

[11] That is, "each decision *has to happen* the way that it does" may be read to imply that agents' mental states are epiphenomenal, or as the fatalistic thesis that "given the past, each decision that happens is inevitable," rather than correctly understood as the deterministic thesis that "necessarily, *given* the past, each decision happens the way it does." Participants may *implicitly* interpret the scope of the modal operator in N&K's description of determinism to entail $[(P \& L) \supset \Box F]$, rather than the proper $\Box[(P \& L) \supset F]$, where P is a description of a prior state of the universe, L the set of the laws of nature, and F a description of a future state of the universe. The latter, but not the former, reading allows that future states *depend* on past states (see Turner and Nahmias, 2006).

[12] For instance, in scenarios presented as descriptions of the real world, 89% of participants said that agents should be held morally responsible and 83% that agents had free will when their decisions were described with the psychological predicates. When the agents were described with the neurobiological, reductionistic predicates, only 40% of participants judged that agents were morally responsible, and only 38% judged that agents had free will.

(e.g., conscious thinking) to be bypassed, but the study did not directly test that hypothesis (see also Shepard, forthcoming).

To investigate whether N&K's (2007) results were confounded by people's confusing determinism with bypassing and whether that might account for some of the difference between their results and NMNT's (2006) earlier findings, we conducted two new studies, presented in the following sections. Only by following Jackson's (1998) advice and controlling for such mistakes can experimental philosophy provide the sort of information about people's reflective intuitions that is needed for philosophical theorizing, such as WRE. Because they do so, we take the results of the studies we present below to provide the best evidence about the ordinary concept of free will currently available.

3. Study 1: Measuring Bypassing

In Study 1, participants were randomly assigned a description of determinism—either N&K's or NMNT's, and in either an abstract or a concrete version, yielding the following four conditions: *N&K abstract, NMNT abstract, N&K concrete,* and *NMNT concrete*.[13] N&K abstract read:

> Imagine a universe (Universe A) in which everything that happens is completely caused by whatever happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe caused what happened next, and so on right up until the present. For example one day John decided to have French Fries at lunch. Like everything else, this decision was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John made his decision, then it *had to happen* that John would decide to have French Fries.
>
> Now imagine a universe (Universe B) in which *almost* everything that happens is completely caused by whatever happened before it. The one exception is human decision making. For example, one day Mary decided to have French Fries at lunch. Since a person's decision in this universe is not completely caused by what happened before it, even if everything in the universe was exactly the same up until Mary made her decision, it *did* ***not*** *have to happen* that Mary would decide to have French Fries. She could have decided to have something different.

[13] Participants were 436 undergraduate students in critical thinking or psychology courses at Georgia State University who completed the entire survey. We excluded 187 participants prior to analysis who (a) responded incorrectly to either of two comprehension questions or (b) completed the survey quickly enough to indicate a lack of attention to the scenario and questions (less than one half of one standard deviation from the mean time for completion), leaving 249 (42% male, 58% female) participants whose data we analyzed. Studies were carried out under previous approval of the University's Institutional Review Board.

> The key difference, then, is that in Universe A every decision is completely caused by what happened before the decision—given the past, each decision *has to happen* the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision *does* ***not*** *have to happen* the way that it does.

In the N&K concrete condition, this scenario was followed by another paragraph:

> In Universe A, a man named Bill has become attracted to his secretary, and he decided that the only way to be with her is to kill his wife and 3 children. He knows that it is impossible to escape from his house in the event of a fire. Before he leaves on a business trip, he sets up a device in his basement that burns down the house and kills his family.

The scenario in the NMNT abstract condition read:

> Imagine there is a universe (Universe C) that is re-created over and over again, starting from the exact same initial conditions and with all the same laws of nature. In this universe the same initial conditions and the same laws of nature cause the exact same events for the entire history of the universe, so that every single time the universe is re-created, everything must happen the exact same way. For instance, in this universe whenever a person decides to do something, *every* time the universe is re-created, that person decides to do the same thing at that time and then does it.

And in the NMNT concrete condition, the last sentence was replaced with:

> For instance, in this universe a person named Jill decides to steal a necklace at a particular time and then steals it, and *every* time the universe is re-created, Jill decides to steal the necklace at that time and then steals it.

After reading one of the four scenarios, participants then indicated their level of agreement to a series of statements on a six-point scale, ranging from ''strongly disagree'' to ''strongly agree.'' These included two sets of questions, each of which was then used to create a composite score for subsequent statistical analyses: an *MR/FW composite score*—an average of participants' answers to three questions about whether an agent in the scenario can be fully morally responsible (MR), can have free will (FW), and deserves to be blamed for their actions; and a *Bypassing composite score*—an average of their answers to four questions about whether the agent's desires, beliefs, and decisions have no effect on what the agent

ends up doing, and whether the agent has no control over what he or she does.[14]

We had several hypotheses. First, we predicted that Bypassing composite scores would be higher in response to N&K's scenarios than NMNT's because of the wording N&K used to describe determinism, and that MR/FW scores would be lower in response to N&K's scenarios than to NMNT's. Second, we predicted that MR/FW scores would be lower, and Bypassing scores higher, for the abstract scenarios compared to the concrete scenarios, because we suspect that descriptions of specific agents and actions are more likely to engage consideration of the causal efficacy of psychological states and processes, such as desires, beliefs, and other reasons-sensitive attitudes, and hence less likely to lead people to think that those states are bypassed. Based on these predictions, we expected that N&K's abstract description of determinism would trigger particularly high Bypassing judgments, explaining why most people in their original study judged that agents could not be "fully morally responsible" in Universe A. These hypotheses were all based on our main prediction: that there would be an inverse correlation between Bypassing and MR/FW judgments across all scenarios. We suspected that the more one interpreted a scenario to involve bypassing, the more one would deny that agents in it could have moral responsibility, free will, and deserve to be blamed, and conversely, that the less one judged a scenario to involve bypassing, the higher one's MR/FW scores would be.

Analyses of the data strongly confirmed each of these predictions (see Figure 1 below and Table 1 in Appendix for descriptive statistics; see Nahmias & Murray 2010 for further discussion of the analyses used in Study 1). MR/FW scores were significantly lower in response to N&K's scenarios than NMNT's, and significantly lower in the abstract

[14] The first statement was always the moral responsibility question (replicating N&K's format), and the remaining statements were randomized to control for order effects. For the exact wording of these statements, see the Appendix and Nahmias and Murray (2010). Responses to the questions used to compute each composite score were highly internally consistent. Reliability analyses produced a Cronbach's alpha of .807 among the questions used to compute the MR/FW composite score and of .823 among the questions used to compute the Bypassing composite score. (Cronbach's alpha is a measure of the interrelatedness and non-uniqueness of the items used to compute it. The more interrelated a series of items and the less unique they are from each other, the higher its value, 1.0 being the highest.) Some might worry that the "no control" question is an inappropriate measure to assess bypassing because they believe that determinism *does* entail that agents have *no* control over what they do. We believe this is mistaken on philosophical grounds. Moreover, removing the "no control" question from the Bypassing composite score actually lowers the Cronbach's alpha among the questions used to compute it from .823 to .797, suggesting that responses to this question are closely related to the responses to the other three bypassing questions.

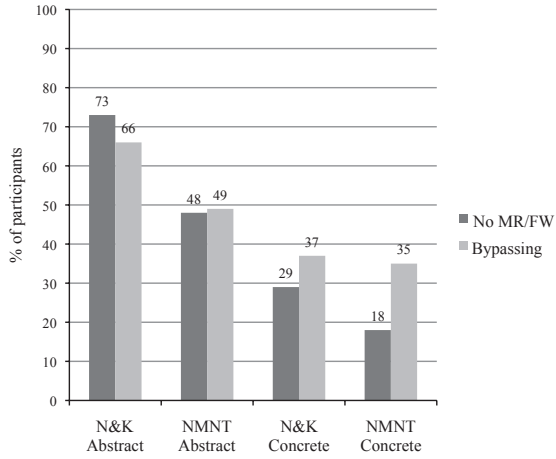

Figure 1. Percentage of participants with MR/FW composite scores < 3.5 midpoint, indicating disagreement on questions about moral responsibility, free will, and blameworthiness (*No MR/FW*) and percentage of participants with Bypassing composite scores > 3.5 midpoint, indicating agreement on questions about bypassing of decisions, desires, beliefs, and control (*Bypassing*). (Data does not include the composite scores for 20 of the 249 participants whose Bypassing composite scores were equal to the 3.5 midpoint score.)

than in the concrete conditions.[15] Bypassing scores were significantly higher in response to N&K's scenarios than in response to NMNT's, and significantly higher in the abstract than in the concrete conditions.[16] As predicted, we also found a very strong inverse correlation between MR/FW and Bypassing scores in every condition (collapsing across all four conditions: $r(247) = -0.734$, $p < .001$).[17]

---

15 A 2 (description: N&K, NMNT) x 2 (condition: abstract, concrete) Analysis of Variance (ANOVA) on the mean MR/FW composite scores showed significant main effects for description: $F(1, 245) = 5.396$, $p = .021$ and for condition: $F(1, 245) = 61.058$, $p < .001$, as well as a marginally significant interaction effect: $F(1, 245) = 3.297$, $p < .071$. An additional pre-planned t-test showed that the mean MR/FW score was significantly lower in N&K abstract than in NMNT abstract: $t(1, 131) = -2.973$, $p = .004$.

16 A 2 (description: N&K, NMNT) x 2 (condition: abstract, concrete) ANOVA on the mean Bypassing composite scores showed a significant main effect for condition: $F(1, 245) = 20.665$, $p < .001$, a near-significant effect for description: $F(1, 245) = 3.463$, $p = .064$, and no significant interaction effect. An additional pre-planned t-test revealed that the mean Bypassing score was significantly higher in N&K abstract than in NMNT abstract: $t(1, 131) = 2.319$, $p = .022$.

17 N&K abstract: $r(75) = -0.695$, $p < .001$; N&K concrete: $r(54) = -0.569$, $p < .001$; NMNT abstract: $r(54) = -0.803$, $p < .001$; NMNT concrete: $r(58) = -0.708$, $p < .001$.

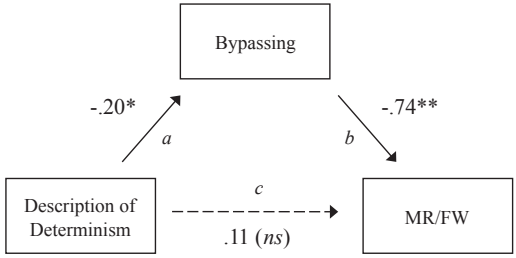

Figure 2. Mediation Analysis. Standardized regression coefficients for the relationships between Description of Determinism (N&K vs. NMNT), MR/FW judgments, and Bypassing judgments. (Sobel test = 2.286, $p < .022$). * $p = .022$, ** $p < .001$.

These results thus confirmed each of our predictions. They also raise a more important question: are the two composite scores merely correlated, or does interpreting a scenario to involve bypassing *cause* participants to judge that agents in it lack moral responsibility and free will? In other words, is interpreting a scenario to involve bypassing what leads people to offer what seem like incompatibilist intuitions? Because if it is, then they aren't actually incompatibilist intuitions at all, since they are not intuitions about the incompatibility of determinism, understood in the way relevant to the philosophical debates, and free will.

To address this causal question, we conducted a mediation analysis using the two abstract scenarios.[18] Mediation analyses involve the specification of a causal model between three variables and the strengths of the paths between them using multiple regression (see Figure 2). They are designed to determine whether the effect of one variable (in our case, the description of determinism: N&K's vs. NMNT's) on another variable (MR/FW judgments) is mediated by a third variable (Bypassing). If the description of determinism no longer has a significant effect on MR/FW once Bypassing is taken into account, then mediation obtains, which provides evidence that the difference between N&K's and NMNT's descriptions of determinism makes a difference to people's judgments about MR/FW *because* of the difference in the degree to which people interpret the descriptions to involve bypassing. And this is exactly what we found. Once we take into account the effect of Bypassing as a mediating variable, the difference between N&K's and NMNT's descriptions of determinism no longer has a significant effect on MR/FW judgments, which suggests that the

[18] Because N&K concrete involves high affect in a way that NMNT concrete does not, we did not include the concrete scenarios in the mediation analysis, as doing so would have introduced a confounding variable: high affect.

difference in MR/FW responses between the two abstract conditions of the scenarios was largely *caused* by people's Bypassing judgments.[19]

This result of the mediation analysis indicates that when participants read a scenario describing determinism, which description they read (N&K versus NMNT) had no significant causal effect on their MR/FW scores over and above the effect it had in virtue of causing different interpretations of whether the scenario involved bypassing. This finding may also provide an alternative explanation of what some (e.g., Knobe and Doris 2010) have taken to be evidence of conflicting intuitions about the relationship between determinism and free will, or of different concepts of free will, in previous experimental studies. Rather than having context-variant intuitions about free will or its relation to determinism (or bypassing), people may simply differ in whether they interpret a description of determinism to involve bypassing, a difference which then causes different responses about agents' free will and responsibility.

These data suggest that most people judge that agents in deterministic scenarios lack moral responsibility, free will, and blameworthiness only when they conflate determinism with bypassing—that is, when they interpret the description of determinism in a scenario to mean that agents' beliefs, desires, and decisions have no effect on what they end up doing and that agents have no control over what they do. When they do not confuse determinism with bypassing, most people do not offer incompatibilist intuitions. Our data thus suggest that many of the participants in N&K's (2007) original study only appeared to have incompatibilist intuitions because they confused determinism with bypassing. These are not genuine incompatibilist intuitions, though, because determinism does not entail, nor is it a form of, bypassing. Compatibilists have long pointed out that determinism does not preclude the causal efficacy of our conscious, rational deliberations and

[19] We conducted three regression analyses to test for mediation, as outlined by Baron and Kenny (1986) and MacKinnon *et al*. (2002). The first regression equation used description of determinism (N&K abstract, NMNT abstract) to predict MR/FW score and yielded a significant effect: $t(132) = 2.973$, $p = .004$, showing that N&K abstract prompted participants to respond that agents do *not* have moral responsibility and free will more than NMNT abstract. The second regression equation estimated changes in Bypassing score using description of determinism and also yielded a significant effect: $t(132) = -2.319$, $p = .022$, showing that participants interpreted N&K abstract to involve bypassing more than they did NMNT abstract. The third equation estimated MR/FW score using both description of determinism and Bypassing score. The link between Bypassing and MR/FW scores was highly significant: $t(132) = -12.799$, $p < .001$, and the relation between description of determinism and MR/FW score was reduced to *non-significance* once Bypassing was included in the model: $t(132) = 1.821$, *ns*; Sobel test = 2.286, $p = .022$. Bypassing accounted for 58% of the total effect of description of determinism on MR/FW score (see Kenny, Kashy, & Bolger 1998, 260–1).

self-control—that it does not entail bypassing. Our results suggest that non-philosophers tend to take determinism to threaten free will and moral responsibility precisely when they fail to see this possibility.

4. Study 2: Controlling for Bypassing

Because many participants in Study 1 did conflate determinism with bypassing, though, we wanted to control for that mistake and *then* assess how many people had compatibilist and incompatibilist intuitions. We also worried that some participants in Study 1 might be making mistakes in the other direction, failing to understand what determinism *does* properly entail—namely, that it *is* impossible, *holding fixed the past and the laws*, for future events to occur otherwise than they actually do. Hence, we wanted to determine if some participants were making that mistake, and if so, to correct for it.[20] Study 2 controlled for these two factors: first, by altering the scenarios to state explicitly that the deterministic universes they describe do not involve bypassing, and second, by excluding participants who misunderstood the modal implications of determinism.[21] By directly manipulating whether the scenarios involved bypassing (relative to Study 1) and measuring the effects of that manipulation, we also aimed to supplement the evidence (from the mediation analysis in Study 1) that bypassing has a *causal* effect on judgments about free will and responsibility.

We predicted that we would replicate the relationships between the Bypassing and MR/FW scores from Study 1, but that telling people that bypassing was not occurring would decrease their Bypassing scores and thereby increase their MR/FW scores. We also predicted that the majority of "competent participants" (as we operationalize the notion)—those who do not conflate determinism with bypassing, but who understand what determinism *does* properly preclude—would give compatibilist responses. In other words, we predicted that when controlling for

[20] Note, however, that in both Studies 1 and 2, we excluded all participants who missed either of two comprehension questions, one of which tested whether they attended to the deterministic language of the scenarios (see Appendix).

[21] Several colleagues have wondered if, by explicitly stating that determinism does not involve or entail bypassing, we lead participants to *do* philosophy, and thereby to no longer provide *pre-theoretical* intuitions. We, too, are wary of "coaching" participants, but eliciting intuitions from people via thought experiments (either experimentally or from the armchair) is already to ask them to do *some* philosophy. And as long as we are trying to elicit people's intuitions, we should attempt to make the meanings of the technical notions we ask them to consider (such as determinism) as clear as possible. We are not interested in people's pre-theoretical understanding of what *determinism* means. The meaning of 'determinism' of the sort relevant to the philosophical debates about free will is not under dispute; what we are interested in is what people's intuitions are about the relationship of *free will* and determinism, where determinism is understood as it is in the philosophical debates.

confusions in people's "first-up" responses as Jackson suggests, their classification of possible cases would reveal the majority to have an underlying concept of free will that does *not* comport with the incompatibilist's conception of free will, 'freedom$_I$'.

4.1. Competent Participants

In Study 2, participants were randomly assigned to read either N&K's or NMNT's abstract scenario, but with a section added to the last paragraph of each that explicitly stated that the scenario did *not* involve bypassing.[22] Thus, N&K abstract in Study 2 included the same first two paragraphs as in Study 1 but then modified the final paragraph to read:

> The key difference, then, is that in Universe A every decision is completely caused by what happened before the decision. This does *not* mean that in Universe A people's mental states (their beliefs, desires, and decisions) have no effect on what they end up doing, and it does *not* mean that people are not part of the causal chains that lead to their actions. Rather, people's mental states *are* part of the causal chains that lead to their actions, though their mental states are always completely caused by earlier things in the causal chain that happened before them—*given* that the past happened the way it did, each decision *has to happen* the way it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision *does not have to happen* the way that it does given what happened in the past.

And NMNT abstract modified the closing sentences to read:

> This does *not* mean that in Universe C people's mental states (their beliefs, desires, and decisions) have no effect on what they end up doing, and it does *not* mean that people are not part of the causal chains that lead to their actions. Rather, people's mental states *are* part of the causal chains that lead to their actions, though their men-

[22] Participants were 302 undergraduate students in critical thinking or psychology courses at Georgia State University who completed the entire survey. We excluded 161 participants prior to analysis who (a) responded incorrectly to either of two comprehension questions or (b) completed the survey quickly enough to indicate a lack of attention to the scenario and questions (less than one half of one standard deviation from the mean time for completion), leaving 141 (39.7% male, 60.3% female) participants whose data we analyzed. Studies were carried out under previous approval of the University's Institutional Review Board. These exclusion criteria were used to maintain consistency with (the identical) criteria used in Study 1 (see note 12), but resulted in a larger percentage of participants being excluded in Study 2 due to completing the survey too quickly. However, if no data are excluded from participants who completed the survey too quickly, the percentages and means do not deviate more than 6% and 0.14, respectively, from those reported in Table 1 (Appendix).

> tal states are always completely caused by earlier things in the causal chain that happened before them—if a person decides to do something in Universe C, then *every* time the universe is re-created with the *same* initial conditions and the *same* laws of nature, that person decides to do the same thing at that time and then does it.

After answering a set of questions about one of these abstract scenarios—including the MR/FW and Bypassing questions—all participants then read the corresponding concrete version of the scenarios, which simply added the following paragraph:

> In Universe [A/C], a man named Bill has become attracted to his coworker and decided that the best way to impress her is to give her an expensive necklace. Bill knows he cannot afford to buy the necklace he wants to give her, but believes he can get away with stealing it from a jewelry store near his home. At a particular time one day (time T), Bill decides to enter the store and steal the necklace and he then does so.[23]

As predicted, the additional language regarding the causal role of agents' mental states led to significantly lower agreement to the Bypassing questions and significantly higher scores for agents' moral responsibility, free will, and blameworthiness in the abstract scenarios (see Table 1 in Appendix for descriptive statistics).[24] For instance, in

---

[23] We thus modified N&K concrete to be low-affect in Study 2, using the same negative action (stealing a necklace) as in NMNT concrete. Two other modifications were also made to Study 2. First, we modified the Bypassing questions, replacing the phrase "...have no effect on what they end up being caused to do" with the more natural-sounding "...have no effect on what they end up doing." We thank Shaun Nichols for this suggestion. Second, as mentioned in the text, all participants in Study 2 first received one of the abstract scenarios, answered the questions about it, and then received the corresponding concrete scenario and responded to it. One other note: in the NMNT concrete scenario and questions that participants actually saw in Study 2, the character's name was 'Frank'. To ease exposition, we have replaced this and used the name 'Bill' for both concrete scenarios in the text and Appendix.

[24] A 2 (study: Study 1, Study 2) x 2 (description: N&K abstract, NMNT abstract) ANOVA on the mean MR/FW composite scores showed significant main effects for study: $F(1, 273) = 5.515$, $p = .020$ and description: $F(1, 273) = 16.723$, $p < .001$, and no significant interaction effect. A 2 (study: Study 1, Study 2) x 2 (description: N&K abstract, NMNT abstract) ANOVA on the mean Bypassing composite scores showed significant main effects for study: $F(1, 273) = 9.656$, $p = .002$ and description: $F(1, 273) = 9.330$, $p = .003$, and no significant interaction effect. The concrete conditions were not included in these analyses because participants in those conditions in Study 2, unlike Study 1, had already responded to the corresponding abstract scenarios. Across all four conditions, reliability analyses produced a Cronbach's alpha of .841 among the questions used to compute the MR/FW composite score, and of .848 among those used to compute the Bypassing composite score.

NMNT abstract, Bypassing judgments went down from 49% to 29%, while MR/FW judgments went up from 52% to 63%.

Again, we found a very strong inverse correlation between Bypassing and MR/FW scores in each of the four conditions (collapsing across all four conditions: $r(292) = -0.724$, $p < .001$).[25] And again, Bypassing score mediated the effect of description of determinism (N&K abstract vs. NMNT abstract) on MR/FW scores.[26] These findings replicate those of Study 1, and they provide further support for our interpretation of them—that mistaking determinism to entail bypassing is a *causal* factor leading participants to offer apparent incompatibilist judgments. We directly intervened on one variable, Bypassing, and it had an effect on the other variable, MR/FW responses. When explicitly informed that bypassing does not occur in a deterministic universe, more participants judge that agents in it can have moral responsibility and free will.[27]

As noted above, one might resist this interpretation due to the possibility that participants do not understand what determinism *does* entail. To address this issue, participants in Study 2 were also asked a "modal question." Those who read N&K abstract, for instance, were asked to rate their agreement with the statement that: "In Universe A, *given* that past events happen the way they do, it *has to happen* that later events happen the way they do."[28] A high proportion of participants answered the modal questions correctly (i.e., agreed): 90% in N&K abstract, 74% in N&K concrete, 81% in NMNT abstract, and 71% in NMNT concrete. Of the subjects who missed the modal questions, the majority

---

25 N&K abstract: $r(66) = -0.618$, $p < .001$; N&K concrete: $r(66) = -0.818$, $p < .001$; NMNT abstract: $r(71) = -0.610$, $p < .001$; NMNT concrete: $r(71) = -0.682$, $p < .001$.

26 This result is, however, marginally significant (Sobel test = 1.945, $p = .052$).

27 Note that in Study 2 N&K abstract explicitly states that the universe it describes does *not* involve bypassing, yet 49% of participants still judged that it *does*. Responses to N&K abstract are also split exactly 50-50% on whether agents in the scenario can be morally responsible, have free will, and deserve blame. We believe these results further suggest that N&K's description of determinism is easily read to involve bypassing, such that when participants are also *told* that it does not involve bypassing, they may simply be unsure about how to interpret the seemingly conflicting information that they are given about the scenario.

28 See Appendix for the "modal question" asked in other scenarios. We removed participants who gave incorrect answers to a comprehension question similar to the "modal question" prior to analysis, helping to explain why many remaining participants answered the question correctly (see notes 13, 22, and Appendix). These modal questions might still be interpreted in ways that misrepresent the proper modal scope of the thesis of determinism, but we believe they adequately balance accuracy and intelligibility. If, as some have suggested, determinism simply *cannot* be properly understood without significant training, then we cannot understand how to interpret the common claims that incompatibilism is intuitive, such as van Inwagen's (2009) above.

had MR/FW scores above the midpoint, indicating agreement. As with apparent incompatibilist responses that are based on conflating determinism with bypassing, these responses are based on a mistaken understanding of determinism, and so should not be taken to reflect intuitions that free will and determinism are compatible.

However, when data from participants who answered the modal question incorrectly or who judged that the scenarios involved bypassing were removed from analysis, most of the remaining participants gave responses indicative of compatibilist intuitions: 62% in N&K abstract, 89% in N&K concrete, 78% in NMNT abstract, and 89% in NMNT concrete.[29] Thus, the percentage of competent participants who did not give incompatibilist responses was comparable across conditions and extremely high (with the exception of N&K abstract; see note 27). We take this convergence to suggest that very few people who understand what determinism does mean actually have incompatibilist intuitions.

This is not to claim that high affect does not bias judgments about moral responsibility and free will, nor to deny that there are interesting differences between abstract and concrete scenarios. Indeed, our results suggest that there is still a difference in attributions of responsibility and free will in abstract vs. concrete scenarios after controlling for level of affect in the latter.[30] Further studies should explore these factors

[29] Data reported on these competent participants does not include the composite scores for 11 of the 141 participants whose Bypassing composite scores were equal to the 3.5 midpoint score.

[30] If we focus only on the competent participants' responses to the concrete scenarios, we see that *9 out of 10* support compatibilism. We are inclined to think that these low-affect concrete cases facilitate comprehension of the underlying philosophical issues better than the abstract scenarios, and so may better elicit the relevant pre-philosophical intuitions, for several reasons. First, scenarios in which specific agents make specific decisions may activate participants' folk psychological thinking (or 'theory of mind' capacities) more than abstract scenarios, thereby leading them to consider agents' mental states more than they do in abstract scenarios. We believe that judgments about moral responsibility and free will are more likely to be reliable and accurate when they do engage these capacities. Assessing whether an agent is morally responsible for something without considering their beliefs, desires, and decisions leaves out crucial information. Second, psychological studies suggest that, in general, reasoning is often more accurate in concrete scenarios (e.g., most participants are able to solve the Wason card selection task when it is made more concrete, whereas most cannot solve it in its original, abstract version (Cox and Griggs 1982)). Third, the reason philosophers develop thought experiments in the first place is typically to elicit people's intuitions based on concrete cases (e.g., Gettier cases), rather than abstract considerations. One of the primary reasons to use the "method of possible cases" is that they can be presented in sufficiently familiar detail to evoke realizations that we may fail to see when considering the relevant philosophical questions more abstractly. As Jackson (1998: 36) says, we consider specific, concrete possible cases to "make explicit what had been implicit in our classificatory practice all along".

(e.g., by using high-affect abstract scenarios), but our results suggest that, *contra* Nichols and Knobe and many incompatibilists, whether elicited in either concrete *or* abstract scenarios, the ordinary concept of free will does not include any commitment to the incompatibility of free will and determinism.[31]

4.2. Choice and the Ability to Do Otherwise

Some may suspect, however, that these so-called ''competent'' participants clearly are not competent enough; that we have failed to apply the method of possible cases with sufficient sophistication, or that the intuitions we have elicited do not represent the reflective judgments that should serve as inputs for theorizing about free will. As Jackson suggests, a concept's theoretical role may carry certain commitments that are not revealed when people are asked directly about that concept. It may be that most people have intuitions about other concepts that are distinct from, but essential components of (or otherwise importantly connected to), their concept of free will, and that their intuitions about these concepts *do* entail that free will is incompatible with determinism (see also Nahmias et al. 2006, 43–45). For instance, perhaps people think free will and responsibility require an unconditional ability to do or choose otherwise—an ability to do or choose otherwise *holding fixed all past conditions and laws*. If so, then even though most participants *respond* that deterministically caused actions are free and blameworthy when asked directly about 'free will' (or 'moral responsibility' or 'blame'), their more basic intuitions about the ability to do or choose otherwise would require $\text{freedom}_{\text{I}}$. Indeed, philosophical arguments for incompatibilism are often attempts to make explicit that our ordinary concepts have just these sorts of implicit commitments (see Sommers 2010).

For instance, Peter van Inwagen's version of the Consequence Argument employs:

---

[31] Our data also speak to previous research by Nichols and Roskies (2008), who suggest that the ordinary concept of free will has a *two-dimensional semantics*, such that people attribute free will to agents in worlds presented as counterfactual (like the scenarios in our studies) based on what they believe free will is in the actual world. According to a two-dimensional account, the extension of 'free will' in worlds considered as counterfactual is determined by the extension of 'free will' in the world considered as actual. If people took 'free will' to refer to $\text{freedom}_{\text{I}}$ in the actual world, then according to a two-dimensional account, they would not attribute free will to agents in possible worlds in which that kind of freedom cannot exist (such as deterministic worlds). That they *do* attribute free will to agents in such worlds in our studies, then, suggests that people either do not believe the actual world is one in which we have $\text{freedom}_{\text{I}}$, or that the concept 'free will' is not two-dimensional (or both).

Rule Beta: Np, N($p \supset q$) $\vdash$ Nq

where 'Np' means "p and no one has, or ever had, any choice about whether p" (van Inwagen 1983, 93). Rule Beta permits transfer of "powerlessness," such that, if no one ever has any choice about whether p, and if no one ever has any choice about whether p implies q, then no one ever has any choice about whether q. From the premises that no one ever has a choice about the actual past and laws of nature, and that no one ever has a choice about our present actions' being the necessary consequences of the past and the laws of nature (as determinism states), Rule Beta then allows us to conclude that, if determinism is true, then no one ever has any choice about their present actions—i.e., that no one has the ability to choose otherwise than they actually do.[32]

Critics of the Consequence Argument often focus on Rule Beta and argue that the conclusion only follows if we interpret the concepts 'ability to do otherwise' or 'choice' *unconditionally*, or categorically. While most compatibilists accept that determinism does entail that it is impossible, *holding fixed the past and laws*, for future events to occur otherwise than they actually do, many deny that this entails that we never have the *ability* to do or choose otherwise than we actually do. According to "classical" compatibilists like Moore and Ayer, for instance, saying that one "could (or has the ability to) do or choose otherwise" just means that "*if* some relevant feature of the past (e.g., one's desires) had been different, *then* one would have chosen or done otherwise," and more recent compatibilists have advanced improved analyses of the ability to do or choose otherwise (e.g., Perry 2004, Vihvelin 2011). In other words, compatibilists often claim that incompatibilists use different conceptions of choice and the ability to do otherwise than those that are relevant to free will and moral responsibility. Incompatibilists, in turn, deny this. Van Inwagen (1983, 105), for instance, suggests that the unconditional conception used in the Consequence Argument is "just that sense of *having a choice* that is relevant in debates about moral responsibility."

[32] Many other incompatibilist arguments employ similar transfer principles. Galen Strawson's Basic Argument (1986), for instance, employs a premise according to which (roughly), an agent cannot be responsible for an action unless she is responsible for (some of) the conditions that bring about that action or the mental states that cause it (cf. the UR principle in Kane 1996). Our argument in the text should apply to such transfer principles in addition to Beta. If these principles are not justified, then arguments including them as premises are unsound.

Thus, the debate again appears to bottom out in a verbal dispute, shifted from the meaning of 'free will' to the meanings of 'choice' and 'ability to do otherwise'. It is relatively uncontroversial that unconditional conceptions of choice and the ability to do otherwise are incompatible with determinism and that conditional or dispositional conceptions are compatible with determinism. The question is which conceptions are relevant, and again, we believe that the philosophical debate should concern those that are used in, and actually matter to, ordinary human practices, especially regarding ascriptions of responsibility. Van Inwagen seems to agree:

> belief in the validity of (Beta) has only two sources, one incommunicable and [the] other inconclusive. The former source is what philosophers are pleased to call "intuition": when I carefully consider (Beta), it seems to be valid... The latter source is the fact that I can think of no instances of (Beta) that have, or could possibly have, true premisses and a false conclusion (1983, 97–98).

Van Inwagen claims that he finds intuitive both Beta itself *and* all particular instances of Beta that he can think of. But if people's responses to possible cases suggest that they believe that agents sometimes have the ability to choose what to do even when their choices are necessary consequences of the past and the laws (which no one has any choice about), then there are numerous intuitive exceptions to particular instances of Beta—namely, typical human choices.[33] What the debate seems to hinge on, then, is what sense of 'choice' and 'ability' ordinary people take to be intuitively relevant to free will and moral responsibility, and again, we believe that an effective way to answer that question is to employ the method of possible cases experimentally.

In Study 2, we decided to investigate how people conceive of choice and the ability to do otherwise in deterministic scenarios by asking participants the following questions (in the abstract and concrete conditions, respectively):

---

[33] Most of the cases van Inwagen considers in "testing" Beta are cases in which it is intuitive that no one has a choice about some outcome, but those intuitions also seem explicable without reference to a principle like Beta. For instance, our intuition that we have no choice about whether life on earth will end in 2100 may not be based on the fact that it is a necessary consequence of the sun's exploding in 2100, which we have no choice about, but simply on the fact that we have no choice about life on earth ending. On the other hand, it is *not* obviously intuitive that I have no choice about whether I now raise my hand *even if* I have no choice about the state of the world a moment earlier and no choice about the fact that the state of the world a moment earlier is sufficient for my hand raising.

Ability: In Universe [A/C], a person has the ability to decide to do something other than what they actually decide to do. (Bill has the ability to decide not to steal the necklace).

Choice: In Universe [A/C], a person has no choice about what they do. (Bill has no choice about what he does).

Among competent participants (those who did not miss the modal question about determinism or mistake the scenarios to involve bypassing), most agreed that agents in the scenarios have the ability to decide otherwise: 58% in N&K concrete, 58% in NMNT abstract, and 70% in NMNT concrete. The only condition in which the majority did not judge that agents have the ability to decide otherwise was N&K abstract, in which only 31% agreed (see note 27). Even more participants judged that agents in the scenarios have a choice about what they do: 59% in N&K abstract, 75% in N&K concrete, 73% in NMNT abstract, and 75% in NMNT concrete. Recall, too, that these are the *competent* participants—those who understand that determinism *does* entail that it *is impossible*, holding fixed the past and the laws, for different future events to occur. For instance, in N&K concrete, 58% of these competent participants judged that Bill "has the *ability* to decide *not* to steal the necklace" and 75% that he "has a *choice*" about doing so, even though they also understood that "*given* everything that happened before time T, Bill *has to* decide to steal the necklace at time T."

Hence, the evidence suggests that most non-philosophers do not share van Inwagen's intuition that the sense of having a choice and the ability to do otherwise relevant to free will and responsibility is *unconditional*. This is not to claim, of course, that most people explicitly have in mind any specific *conditional* analysis of the ability to do or choose otherwise. Many people may simply have an implicit understanding of contingent events in general, including human decisions, such that they could have happened otherwise only if something leading up to them had happened otherwise.[34] Indeed, this way of understanding counterfactuals is quite natural. When we believe that the storm might have passed, that the dropped plate might not have shattered, or that the dog could have caught the frisbee, we do not seem to mean that these

[34] It may also be that some of the competent participants in the minority who are willing to attribute free will and responsibility to agents, but who also judge that the agents have no choice about what they do and/or lack the ability to decide otherwise, are expressing semi-compatibilist intuitions (i.e., the belief that one can be responsible even if one is not able to do otherwise; Fischer 1994). Some of the participants might also be thinking about choice and the ability to do otherwise in terms of *general capacities* of the agent rather than *specific abilities* to do otherwise at the particular time in question (see Campbell 1997).

alternative events could have occurred *holding fixed everything about the past and laws of nature*. How could the dog have caught the frisbee holding fixed the fact that he jumped too short, failed to open his mouth quite in time, and the like? Rather, we seem to believe that slightly different past conditions could have led to these different outcomes—different meteorological conditions, the plate's falling at a different angle, the dog's jumping a bit higher. Whether people treat humans' ability to do otherwise as metaphysically different in kind from other contingent events is an open question, but our data offer initial evidence that the ordinary understanding of human choice and action is *not* unconditional. Whatever concepts of ability and choice people use in thinking about free will and moral responsibility, most do not take them to be threatened by determinism, so long as they do not conflate determinism with bypassing.

Our data thus provide some evidence that the Consequence Argument and related incompatibilist arguments (see note 32) are either unsound or irrelevant. There certainly are conceptions of an unconditional ability to do or choose otherwise, holding fixed the actual past and the laws, according to which the argument is sound. But our results suggest that these may be technical notions and *not* the concepts of the ability to do and choose otherwise that most people actually employ when making judgments about free will and responsibility. Thus, if an incompatibilist argument employs an unconditional reading, then its conclusion, while true, is that we cannot have a kind of ability that most people do not talk, think, or care about. If the argument uses a conditional concept of ability or choice instead, then typical human choices serve as intuitive counterexamples to Beta (or whatever related transfer principle the argument employs), and the argument thereby fails.[35]

5. Philosophical Implications

We find it unlikely that most people have any explicit theory of choice or the ability to do otherwise that contains substantive metaphysical commitments, just as we find it unlikely that most people have any explicit theory about the relation between free will and determinism.

[35] Another possibility is that people find Beta to be intuitive in its abstract formulation, but in the process of WRE, would take it to be less intuitive than counterexamples to Beta involving concrete cases of human choices. In that case, the incompatibilist would have to provide arguments for why any intuitive support for the abstract principle should outweigh the stronger intuitive support for concrete counterexamples to it. Even van Inwagen (1983) suggests that, in WRE, if he had evidence that determinism was true, he would reject Beta rather than rejecting judgments that humans can be morally responsible for their actions.

This raises a point worth stressing. We have argued that most people do not have incompatibilist intuitions, but this is not to claim that they *do* have intuitions or theories with specific compatibilist content. If intuitions are mental states with representational content, then we suspect that there are no *bona fide* incompatibilist *or* compatibilist intuitions at all. Most non-philosophers have probably never contemplated determinism, let alone formed any mental representations *that* it is compatible or incompatible with free will. The question, though, is whether the mental states that people *do have* (their intuitions, judgments, beliefs, concepts, theories, etc.) about free will, or about conditions required for free will or responsibility, represent it in a way that is *consistent* with compatibilism. Our results suggest that they do. Whatever thoughts people have about free will, their content does *not* seem to commit them to its incompatibility with determinism, contra many incompatibilists' claims and Nichols and Knobe's (2007) interpretation of their data.

We suggested at the outset that the debate between compatibilists and incompatibilists depends crucially on determining what people mean by 'free will', since there may be little substantive dispute about the compatibility question once the question of meaning is resolved. If $\text{freedom}_{C,}$ and not $\text{freedom}_{I,}$ is what ordinary people care about having and worry about not having (and if no other closely related concepts commit people to incompatibilism, either), then why should we accept a theory of free will that refers to freedom_{I}? We also argued that the ordinary concept 'free will' could be fruitfully studied experimentally. Following Jackson's (1998) advice, we employed the method of possible cases in ways that controlled and corrected for several of the most common classificatory mistakes that people might make.[36]

[36] Perhaps some of the participants we classified as competent made other mistakes that, if corrected for, would increase the estimated number of those with incompatibilist commitments. Some may have failed to understand the "modal restrictions" imposed by determinism, despite our attempts to control for such confusion. Or, it may be that 'free will' plays theoretical roles (other than those relating to 'choice' and the 'ability to do otherwise') that we did not investigate but which require its referent to be incompatible with determinism. Though we asked about deserving blame, perhaps people take retributive punishment to require "ultimate sourcehood." While these possibilities deserve further investigation, we think the prediction that most people will still turn out to be intuitive incompatibilists is a bad bet. After all, the studies presented here only attempted to control for a subset of the ways in which determinism might be misunderstood in the other direction—those that involve some (but by no means all) types of bypassing. If some of the participants that we have classified as competent still conflated determinism with other threatening theses, such as coercion or manipulation, or if they implicitly anthropomorphized the laws of nature or causation by past events, then it may be that even fewer *truly* competent participants have incompatibilist intuitions. Our best current estimate is that very few do.

Our results demonstrate that the appearance of widespread incompatibilist intuitions among the folk is an illusion, based on the common confusion between determinism and bypassing. Incompatibilism as a philosophical thesis, then, is likely either false, or else the uninteresting thesis that determinism is incompatible with some kind of "free will" that most non-philosophers do not seem to talk, think, or care about.

Our results also raise an important further question—namely, *why* determinism and bypassing are so easily conflated, and what the fact that they are might tell us about our ordinary understanding of free will.[37] In some cases, determinism may simply be introduced in a misleading way. But the ease with which people take determinism to entail bypassing may also be due to an intuitive aversion to causal overdetermination. Some participants in our studies may have taken the presence of factors external to agents' psychologies (long-past events and the laws of nature) that were *sufficient* for agents' actions to imply that those agents' psychologies were not part of the causal chains that led to their actions. And perhaps this is just one instance of a more general tendency to assume (presumably implicitly) that a complete causal explanation of a phenomenon Z in terms of X precludes any other causal contribution to Z from other sources.

On the "horizontal," or temporal, dimension, this *single explanation assumption* (SEA) might take the form: if event X at time t_1 is causally sufficient for event Z at t_3, then no event Y at t_2 plays any causal role in bringing about Z. In this form, SEA is clearly misleading, since X might bring about Z *by causing* Y to cause Z (e.g., when lighting a fuse makes a bomb explode *by* causing the fuse's burning to detonate the bomb). Similarly, if determinism is true, long-past conditions are sufficient to bring about our actions, but perhaps only by bringing about intermediate mental states that in turn cause those actions.

On the "vertical," constitution or supervenience, dimension, SEA might take a form reminiscent of Jaegwon Kim's causal exclusion argument (1998): if Y supervenes on X (such that X is sufficient for Y) and if X is causally sufficient for Z, then Y is causally irrelevant to Z (or at best an overdetermining cause of Z). If people interpret determinism as a type of reductionism, according to which there are lower-level, non-mental events sufficient to bring about their actions, then they might take determinism to imply that there is no room left for their mental activity or conscious selves to do any causal work in bringing about their actions.

On either the horizontal or vertical dimension of SEA, people might take determinism to entail bypassing even if, in fact, the past and laws

[37] We thank Joshua Knobe for pressing this question, though not the specific suggestion we make in the text.

of nature must "go through" one's mental life to produce one's present actions, and even if lower-level (e.g., neurobiological) causes do not pre-empt the causal relevance of the higher-level mental processes that supervene on, or are identical to, them. If the latter is what drives most people's worries about whether we have free will and moral responsibility, then the *real* problem of free will may be just one instance of the mind-body problem (cf. Davidson 1980, 207).

As easily as they might be confused, however, whether causal exclusion threatens free will is a separate question from whether determinism does. Determinism does not entail that the mental supervenes on, or reduces to, the physical (nor do physicalism or reductionism entail determinism). Our results suggest that most people do *not* take determinism to preclude free will. But Nahmias et al. (2007) provide evidence that most people *do* take free will to be threatened by the complete causation of actions by neurobiological states (cf. Shepard, forthcoming), and our present results suggest that bypassing of mental states would be a genuinely intuitive threat to free will, as well. Given that a growing number of scientists claim that their results show that our conscious mental states often *are* bypassed (e.g., Wegner 2002 and Bargh 2008; cf. Nahmias, forthcoming), philosophers working on free will should shift their attention. Our primary aim in this paper has been to address what people mean by 'free will'. Our answer suggests that to address whether we actually have free will, energy would be better spent on investigating the merits of these potential threats from the philosophy and sciences of the mind, rather than on assessing what its relation is to the empty threat of determinism.[38]

Appendix

MR/FW questions (abstract cases first with variations in brackets; concrete cases in parentheses)

MR: In Universe [A/C], it is possible for a person to be fully morally responsible for their actions. ([Bill/Jill] is fully morally responsible for [killing his wife and children/stealing the necklace].)

[38] Earlier versions of this paper were presented at the 2010 meeting of the Society for Philosophy and Psychology; the Metro Experimental Research Group (MERG), February 2011; and to audiences at the University of California, Berkeley. We thank the audiences on those occasions for their helpful feedback. For valuable comments on various stages of this work, we especially thank Brian Berkey, David Blumenfeld, Fiery Cushman, Dan Dennett, Peter Epstein, John Martin Fischer, George Graham, Joshua Knobe, Trevor Kvaran, Neil Levy, Tania Lombrozo, Al Mele, Stephen Morris, Thomas Nadelhoffer, Shaun Nichols, Jonathan Phillips, Jason Shepard, Tamler Sommers, Chandra Sripada, Reuben Stern, Bradley Thomas, Jonathan Weinberg, and Dan Weiskopf.

FW: In Universe [A/C], it is possible for a person to have free will. (It is possible for [Bill/Jill] to have free will.)

Blame: In Universe [A/C], a person deserves to be blamed for the bad things they do. ([Bill/Jill] deserves to be blamed for [killing his wife and children/stealing the necklace.])

Bypassing questions (as worded in Study 2; abstract cases first with variations in brackets; concrete cases in parentheses)

Decisions: In Universe [A/C], a person's decisions have no effect on what they end up doing. (Bill's decision to steal the necklace has no effect on what he ends up doing.)

Wants: In Universe [A/C], what a person wants has no effect on what they end up doing. (What Bill wants has no effect on what he ends up doing.)

Believes: In Universe [A/C], what a person believes has no effect on what they end up doing. (What Bill believes has no effect on what he ends up doing.)

No Control: In Universe [A/C], a person has no control over what they do. (Bill has no control over what he does.)

Modal questions (Study 2)

N&K Abstract: In Universe A, *given* that past events happen the way they do, it *has to happen* that later events happen the way they do.

N&K Concrete: *Given* everything that happened before time T, Bill *has to* decide to steal the necklace at time T.

NMNT Abstract: In Universe C, if the universe is re-created with the exact *same* initial conditions and laws of nature, then it *has to happen* that later events happen the way they do.

NMNT Concrete: If Universe C were re-created with the exact *same* initial conditions and laws of nature, it is possible Bill would *not* decide to steal the necklace at time T. (reverse scored)

Ability and Choice questions (Study 2)

Ability Abstract: In Universe [A/C], a person has the ability to decide to do something other than what they actually decide to do.

Ability Concrete: Bill has the ability to decide not to steal the necklace.

Choice Abstract: In Universe [A/C], a person has no choice about what they do. (reverse scored)

Choice Concrete: Bill has no choice about what he does. (reverse scored)

Sample Comprehension questions

N&K Abstract: According to the scenario, in Universe A, everything that happens is completely caused by what happened before it.

NMNT Abstract: According to the scenario, in Universe C the same initial conditions and the same laws of nature cause the exact same events for the entire history of the universe.

Table 1.

MR/FW and Bypassing Descriptive Statistics.

Means (on six-point scale), percentages of participants with scores > 3.5 midpoint score (i.e., expressing agreement), and standard deviations for MR/FW and Bypassing composite scores for each scenario in Studies 1 and 2.

				MR/FW			Bypassing		
Description	Condition	Study	N	Mean	%	S.D.	Mean	%	S.D.
N&K	Abstract	1	77	2.818	27	1.204	3.958	66	1.205
		2	68	3.221	50	1.178	3.434	49	1.308
	Concrete	1	56	4.363	71	1.298	3.018	37	1.216
		2	68	4.069	72	1.348	2.886	25	1.225
NMNT	Abstract	1	56	3.482	52	1.360	3.442	49	1.346
		2	73	3.785	63	1.216	3.014	29	1.197
	Concrete	1	60	4.444	82	1.174	2.946	35	1.183
		2	73	4.598	81	1.202	2.425	12	1.092

References

Alexander, J., Mallon, R. and Weinberg, J. 2010. Accentuate the negative. *Review of Philosophy and Psychology*, 1: 297–314.

Bargh, J. 2008. Free will is un-natural. In J. Baer, J. Kaufmann and R. Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 128–154). New York: Oxford University Press.

Baron, R. M. and Kenny, D. A. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51: 1173–1182.

Blumenfeld, D. 1988. Freedom and mind control. *American Philosophical Quarterly*, 25: 215–227.

Campbell, J. K. 1997. A compatibilist theory of alternative possibilities. *Philosophical Studies*, 88: 319–330.

Chalmers, D. J. 2011. Verbal disputes. *Philosophical Review*, 120(4): 515–566.

Cover, J. A. and O'Leary-Hawthorne, J. 1996. Free agency and materialism. In J. Jordan and D. Howard-Snyder (Eds.), *Faith, freedom and rationality* (pp. 47–71). Lanham, MD: Roman and Littlefield.

Cox, J. R. and Griggs, R. A. 1982. The effects of experience on performance in Wason's selection task. *Memory and Cognition*, 105: 496–502.

Daniels, N. 1979. Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy*, 76: 256–282.

Davidson, D. 1980. Mental events. In *Essays on actions and events* (pp. 207–227). New York: Oxford University Press.

Ekstrom, L. 2002. Libertarianism and frankfurt-style cases. In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 309–322). New York: Oxford University Press.

Fischer, J. M. 1994. *The metaphysics of free will*. Oxford: Blackwell Publishers.

—— and Ravizza, M. 1998. *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.

Frankfurt, H. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy*, 68: 5–20.

Goodman, N. 1955. *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Harris, S. 2012. *Free will*. New York: Free Press.

Kane, R. 1996. *The significance of free will*. New York: Oxford University Press.

—— 1999. Responsibility, luck, and chance: reflections on free will and indeterminism. *Journal of Philosophy*, 96: 217–240.

—— 2005. *A contemporary introduction to free will*. New York: Oxford University Press.

Kenny, D. A., Kashy, D. A. and Bolger, N. 1998. Data analysis in social psychology. In D. Gilbert, S. Fiske and G. Lindzey (Eds.), *The handbook of social psychology: Vol. 1*, 4th ed. (pp. 233–265). Boston: McGraw-Hill.

Knobe, J. and Nichols, S. 2008. An experimental philosophy manifesto. In J. Knobe and S. Nichols (Eds.), *Experimental philosophy* (pp. 3–14). New York: Oxford University Press.

—— and Doris, J. 2010. Responsibility. In J. Doris (Ed.), *The moral psychology handbook* (pp. 321–353). New York: Oxford University Press.

Lombrozo, T. 2009. The role of moral commitments in moral judgment. *Cognitive Science*, 33: 273–286.

Mele, A. 2005. *Autonomous agents*. New York: Oxford University Press.

Monroe, A. E. and Malle, B. F. 2009. From uncaused will to conscious choice: The need to study, not speculate about people's folk concept of free will. *Review of Philosophy and Psychology*, 1: 211–224.

Nadelhoffer, T. and Nahmias, E. 2007. The past and future of experimental philosophy. *Philosophical Explorations*, 10: 123–149.

Nahmias, E., Morris, S., Nadelhoffer, T. and Turner, J. 2005. Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, 18: 561–584.

——, ——, —— and —— 2006. Is incompatibilism intuitive? *Philosophy and Phenomenological Research*, 73: 28–53.

——, Coates, D. J. and Kvaran, T. 2007. Free will, moral Responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy*, 31: 214–242.
—— and Murray, D. 2010. Experimental philosophy on free will: An error theory for incompatibilist intuitions. In J. Aguilar, A. Buckareff and K. Frankish. (Eds.), *New Waves in Philosophy of Action*, (pp. 189–215). New York: Palgrave-Macmillan.
—— Forthcoming. Is free will an illusion? Confronting challenges from the modern mind sciences. In W. Sinnott-Armstrong. (Ed.) *Moral Psychology, vol. 4, Free Will and Responsibility*, New York: Oxford University Press.
Nichols, S. and Knobe, J. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs*, 41: 663–685.
—— and Roskies, A. 2008. Bringing moral responsibility down to earth. *Journal of Philosophy*, 105: 371–388.
—— 2011. Experimental philosophy and the problem of free will. *Science* 331(6023): 1401–1403.
O'Connor, T. 2000. *Persons and causes: The metaphysics of free will*. New York: Oxford University Press.
O'Keefe, T. 2005. *Epicurus on freedom*. Cambridge: Cambridge University Press.
Pereboom, D. 2001. *Living without free will*. Cambridge: Cambridge University Press.
Perry, J. 2004. Compatibilist options. In D. Shier, M. O'Rourke and J. K. Campbell (Eds.), *Freedom and determinism* (pp. 231–254). Cambridge, MA: MIT Press.
Pink, T. 2004. *Free will: A very short introduction*. New York: Oxford University Press.
Rawls, J. 1971. *A theory of justice*. Cambridge, MA: Harvard University Press.
Schlick, M. 1939. *Problems of ethics*, ed. by D. Rynin. New York: Prentice Hall.
Shepard, J. Forthcoming. Free will and consciousness: Experimental studies. *Consciousness and Cognition*.
Sommers, T. 2010. Experimental philosophy and free will. *Philosophy Compass*, 5: 199–212.
Sripada, C. and Konrath, S. 2011. Telling more than we can know about intentional action. *Mind & Language*, 26: 353–380.
—— Forthcoming. What makes a manipulated agent unfree? *Philosophy and Phenomenological Research*.
Stich, S. 1988. Reflective equilibrium: Analytic epistemology and the problem of cognitive diversity. *Synthese*, 74: 391–413. Reprinted,

with minor changes, in M. DePaul & W. Ramsey (Eds.), *Rethinking intuition* (pp. 95-112). Lanham, MD: Rowman & Littlefield.
—— and Weinberg, J. 2001. Jackson's empirical assumptions. *Philosophy and Phenomenological Research*, 62: 637–643.
Strawson, G. 1986. *Freedom and belief*. Oxford: Clarendon Press.
Strawson, P. 1962. Freedom and resentment. Reprinted in G. Watson (Ed.), *Free will*, 2nd ed. (pp. 72–93). New York: Oxford University Press.
Swanton, C. 1992. *Freedom: A coherence account*. Indianapolis, IN: Hackett Publishing.
van Inwagen, P. 1983. *An essay on free will*. Oxford: Clarendon Press.
—— 2009. *Metaphysics*, 3rd ed. Boulder, Co: Westview Press.
Vargas, M. 2004. Responsibility and the aims of theory: Strawson and revisionism. *Pacific Philosophical Quarterly*, 85: 218–241.
—— 2009. Revisionism about free will: A statement & defense. *Philosophical Studies*, 144: 45–62.
—— 2010. The revisionist turn: A brief history of recent work on free will. In J. Agúilar, A. Buckareff and K. Frankish (Eds.), *New Waves in Philosophy of Action*. New York: Palgrave-Macmillan.
Vihvelin, K. 2011. How to think about the free will/determinism problem. In J. Campbell and M. O'Rourke (Eds.), *Carving nature at its joints*. (pp. 314–340). Cambridge, MA: MIT Press.
Wallace, R. J. 1994. *Responsibility and the moral sentiments*. Cambridge, MA: Harvard University Press.
Wegner, D. 2002. *The illusion of conscious will*. Cambridge, MA: MIT Press.
Weinberg, J., Nichols, S. and Stich, S. 2001. Normativity and epistemic intuitions. *Philosophical Topics*, 29: 429–460.
—— 2007. How to challenge intuitions empirically without risking scepticism. *Midwest Studies in Philosophy*, 31: 318–343.