



Experimental Philosophy and the Problem of Free Will

Shaun Nichols

Science **331**, 1401 (2011);

DOI: 10.1126/science.1192931

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 28, 2011):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/331/6023/1401.full.html>

This article **cites 18 articles**, 1 of which can be accessed free:

<http://www.sciencemag.org/content/331/6023/1401.full.html#ref-list-1>

This article appears in the following **subject collections**:

Psychology

<http://www.sciencemag.org/cgi/collection/psychology>

Downloaded from www.sciencemag.org on April 28, 2011

Experimental Philosophy and the Problem of Free Will

Shaun Nichols

Many philosophical problems are rooted in everyday thought, and experimental philosophy uses social scientific techniques to study the psychological underpinnings of such problems. In the case of free will, research suggests that people in a diverse range of cultures reject determinism, but people give conflicting responses on whether determinism would undermine moral responsibility. When presented with abstract questions, people tend to maintain that determinism would undermine responsibility, but when presented with concrete cases of wrongdoing, people tend to say that determinism is consistent with moral responsibility. It remains unclear why people reject determinism and what drives people's conflicted attitudes about responsibility. Experimental philosophy aims to address these issues and thereby illuminate the philosophical problem of free will.

Many central philosophical problems—such as problems concerning free will, morality, and consciousness—have their roots in our ordinary ways of understanding the world. It takes no special training to come to appreciate questions like “How can a material object be conscious?” or “Is morality only relative to culture?” Such problems resonate with common sense. These common-sense philosophical problems are also notorious for their resilience. Many of them stretch back to the earliest days of philosophy. The problems have persisted, and to a great extent, so have the techniques used to advance the inquiry. The central technique is careful and sustained thought, sharpened by dialogue with fellow philosophers.

Experimental philosophy is a recent movement that brings new techniques to bear on philosophical problems (1). The techniques are actually familiar experimental methods from social science, but the targets of these experiments are distinctively philosophical, like the common-sense philosophical problems surrounding free will, morality, and consciousness. Because these problems are grounded in common sense, experimental philosophers aim to diagnose the psychological origins of philosophical problems, and such diagnoses might indicate new avenues of treatment.

In some cases, a common-sense philosophical problem arises because it is difficult for us to make sense of the way the world works. For instance, it is difficult to understand, intuitively, how a physical thing like the brain can produce the conscious experience of tasting cantaloupe, which seems utterly different from anything physical (2). In other cases, common-sense philosophical problems arise because we seem to have conflicting intuitions. Here, free will provides a vivid example. On the one hand, it seems that, at the moment of making a decision about whether to stop writing (for instance), it is genuinely possible for one to freely decide either way, and

“... experimental philosophers aim to diagnose the psychological origins of philosophical problems.”

after the decision is made, it still seems that one could have decided something else, even if everything leading up to the decision had been exactly the same (3, 4). On the other hand, it also seems intuitive that when an event occurs, there should be a complete causal explanation for why it happened. Every event seems to be caused by the preceding events. It would be strange to say that the choice to continue writing just happened, with no causal explanation. The idea that everything that happens must have a causal explanation is central to determinism, the view that (roughly) every event is completely caused by what happened leading up to the event. Determinism is intuitively attractive, but it seems to conflict with the idea that, at the moment of a decision, it is genuinely possible to choose one way or the other (5, 6). It seems like something has to give, either our commitment to free will or our commitment to the idea that every event is completely caused by the preceding events.

Agency and Determinism

Philosophers have explored the apparent intuitive conflict between free will and determinism by invoking sophisticated logical arguments and semantic analyses (7, 8). Experimental philosophers focus instead on the psychological aspects of the problem. Are people really pulled in these

different directions? Do ordinary people really have intuitions at odds with determinism, as philosophers suggest? And if so, what is the ultimate psychological source of these intuitions?

A person typically will not regard her current action as free unless he or she feels like it is her own voluntary action. Researchers have proposed two very different factors that contribute to this sense of agency. According to one hypothesis, the internal motoric signals that cause behavior also generate a prediction about imminent bodily movement, and this prediction is compared to the actual sensory information of bodily motion. If the predicted movement conforms to the sensory information, then one gets the feeling of agency; otherwise the movement is likely to feel involuntary (9). Another proposal maintains that a person's sense of agency is affected by the presence of external cues, such as the time interval between an external cue and a subsequent behavior (10). These internal and external cues both seem to contribute to the sense of agency (11).

The sense of agency is critical to any sense that one's current action is free, because an action will not feel free unless it feels like one's own action. But there is a further question about whether one's actions are regarded as undetermined. Developmental psychologists have recently shown that, in some circumstances, young children reason in a way that suggests a belief in determinism.

When observing physical events like a light going on, children expect there to be a causal explanation for the event. After they have observed a light going on when a switch is flipped, if the switch subsequently fails to turn on the light, young children reliably search for a causal explanation for why the light did not turn on (12). However, experimental philosophers have also found evidence suggesting that 3- to 5-year-old children reject determinism in the context of human action. With a child observing, an experimenter performed a simple action such as putting her hand in a box, and the child was then asked whether the experimenter could have done something else. The vast majority of children said that the person could have done something else (13, 14). Most children did not, however, say the same thing after observing a physical event like a ball rolling into the box. Rather, in that case, children denied that the ball could have done something else. The conflicting intuitions that give rise to the problem of free will may already be present at an early age.

A different way to assess whether people reject determinism is to present them with a non-technical description of a deterministic universe and then gauge their reaction. This method has now been employed in several studies on adults. In one study, a deterministic universe was char-

acterized as follows: “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” (15). After reading such a description, adult participants in the U.S. tended to reject the idea that our universe is like this, at least when it comes to human decision-making. Subsequent cross-cultural research finds the same pattern in Chinese, Indian, and Colombian populations (16).

These results from experimental philosophy confirm what many philosophers already maintained: that common sense is committed to indeterminism about decision-making (3–6). But the results underscore a puzzling aspect of this common-sense commitment. What leads people to reject statements of determinism? What are the psychological sources of this reaction? After all, determinism is a sophisticated theory of the universe. And it is likely that many of the participants in these experiments had never previously given determinism much thought. Thus, one goal of recent work has been to figure out why people tend to converge on the rejection of determinism.

One explanation with an impressive philosophical pedigree is that people reject determinism because introspection does not reveal a deterministic set of causes of action (3, 17, 18). Often when we introspect on our reasons for performing an action, the reasons we perceive do not univocally point to a particular action. Cognitive scientists widely agree that introspection fails to reveal all of the causal influences on our actions (19, 20). But it remains unclear whether this introspective limitation provides a complete explanation for why people reject determinism.

An alternative proposal focuses on the way people think about the self as an agent (21, 22). As noted above, research on the sense of agency shows that both internal and external cues contribute to the sense that “I” produced an action (9–11). But how do people think of the self that produces the action? Determinism might be counterintuitive because of the way people think of the self (23). A computer’s behavior is obviously a function of the computer program and the inputs. But the self, accord-

ing to this proposal, is conceived as something distinct from psychological mechanisms and inputs, something that can stand apart from those details and make an executive decision. If people think of the self as something apart from the inputs and mechanisms, then it is not surprising that they reject the idea that decisions are produced by deterministic mechanisms and processes.

Moral Responsibility and Determinism

Suppose it turns out that determinism is true. What does that mean in terms of whether people are morally responsible for their actions? Philosophers have long been divided on this question. One view, compatibilism, maintains that determinism is quite irrelevant to the issue of moral responsibility; i.e., our ordinary ways of thinking about responsibility are immune to any threat from determinism (24). The opposing view, incompatibilism, maintains that if determinism is true, no one is truly morally responsible (4, 5, 25).

Experimental philosophers have shown that ordinary people, like philosophers, offer conflicting views. In one experiment, participants were

presented with a description of a deterministic universe in which “Every decision is completely caused by what happened before the decision—given the past, each decision has to happen the way that it does.” After being presented with this description of determinism, one group of participants was asked whether it is possible for anyone to be morally responsible for their actions in such a universe. These participants tended to say that it is not possible to be morally responsible in that universe. That question about moral responsibility is, of course, pitched at an abstract level. Another group of participants was presented instead with a concrete case of a man who killed his family. That provoked a much different response. When presented with a concrete case of man performing a reprehensible action, people tended to say that the man was fully morally responsible for his actions, even when set in a deterministic universe (Fig. 1A) (15). Indeed, concrete cases of bad behavior lead people to attribute responsibility, even when the action is caused by a neurological disorder (26). In contrast, when asked abstract questions about whether people in a deterministic universe should be blamed

for bad actions, people tend to say that they should not be blamed. Moreover, a study on a diverse set of cultures—including the U.S., China, and India—found no cultural differences on these kinds of questions. People in all cultures tended to maintain that in a deterministic universe, people would not be morally responsible for their actions (16).

Related results emerge when participants are presented with the vivid possibility that our own universe is deterministic. In one condition, participants are told that many eminent scientists are convinced that determinism is true in our world: “that every decision a person makes is completely caused by what happened before the decision—given the past, each decision has to happen the way that it does.” In the other condition, participants are told to imagine another universe in which many eminent scientists are convinced that determinism is true. In both conditions, they are then asked to consider whether, if those scientists are right, people are morally responsible. When imagining alternate universes, participants tend to give incompatibilist responses, saying that people are not morally responsible in that case. However, when led to consider our own universe as deterministic, participants were more likely to say that people would still be morally responsible (27, 28). Imagining a less distant scenario makes people give more compatibilist responses.

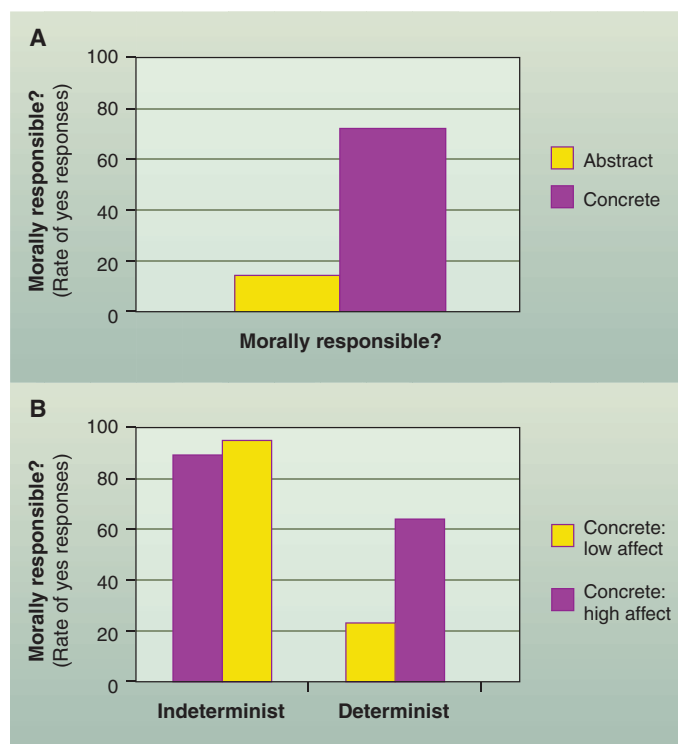


Fig. 1. (A) After reading a nontechnical description of determinism, one group of participants was asked an abstract question about whether it is possible to be morally responsible in such a universe. The other group was given a concrete scenario in which a particular person did something terrible in that universe, and they were asked whether that person is morally responsible. **(B)** After reading the description of determinism, participants were presented with either a concrete case of an emotionally upsetting (high-affect) transgression or a concrete case of an emotionally bland (low-affect) transgression. In half of the cases, the transgression was set in an indeterminist universe. In that case, participants were equally likely to attribute responsibility in both high- and low-affect conditions. In the other half of the cases, the transgression occurred in a determinist universe. In that case, participants were significantly more likely to attribute responsibility in the high-affect condition (12).

What explains these conflicting responses? One class of explanation holds that the responses in the abstract condition reveal the true common-sense view. When they are calm and collected, people acknowledge that determinism precludes responsibility. In contrast, when faced with cases that are concrete and close to home, our responses are biased by emotional and motivational factors. Our desire to blame the offender (29) drives us to blame, even under the assumption of determinism (15). Previous research indicates that emotional activation does affect attribution of blame; i.e., participants who have been made to feel negative emotion issue harsher judgments of blame (30). Additionally, there is reason to suspect that some of the manipulations in the free-will experiments do differentially trigger affect. For instance, concrete descriptions of reprehensible actions are likely to be more upsetting than abstract questions about moral responsibility. This might explain why people in the concrete condition are more likely to attribute responsibility; their emotions are driving their judgments of responsibility. Some evidence in favor of this interpretation comes from the fact that when participants in the determinist experiments are presented with an emotionally charged “high-affect” concrete transgression (i.e., rape), they are much more likely to attribute responsibility than when they are presented with a low-affect concrete transgression (i.e., tax evasion) (Fig. 1B) (15).

A very different explanation of people’s conflicting judgments focuses on the incompatibilist responses and maintains that these responses result from a misunderstanding about the nature of determinism (31, 32). The hypothesis maintains that if people think the world is deterministic, their conscious deliberations and other psychological processes are causally irrelevant. However, this is not what determinism entails. Determinism is consistent with the idea that behavior is produced (i.e., determined) by conscious psychological processes. To explore whether this kind of confusion affects responses, participants were presented with different versions of determinism. In one condition, behavior was said to be determined by neurological and chemical processes. In that case, participants drew incompatibilist conclusions, saying that people were not responsible if behavior is all caused by neurological and chemical processes. In the other condition, behavior was characterized as being determined by psychological processes. In that case, participants tended to give compatibilist responses, maintaining that people are responsible even if their actions are determined by their psychological states (31). This suggests that determinism itself might be less threatening to our ordinary ideas of moral responsibility, so long as the determining causes are our psychological states and processes.

Potential Implications

Just as philosophers have advocated diametrically opposed views on whether determinism

undermines responsibility, ordinary people offer conflicting views on the matter. In some cases, people tend to say that determinism would undermine responsibility; in other cases, they tend to regard determinism as compatible with responsibility. Experimental philosophy holds out the promise of explaining why we are pulled in different directions. The extant work suggests that people are pulled in different directions because different mental mechanisms are implicated in different conditions. As noted above, one hypothesis is that incompatibilist responses—that responsibility is not compatible with determinism—are favored by mental mechanisms devoted to unemotional abstract reasoning. In contrast, compatibilist views seem to be favored when emotional attitudes are triggered by concrete scenarios. If something along these lines is right, then this would help explain the persistence of the philosophical dispute over free will and moral responsibility. Part of the reason that the problem of free will is so resilient is that each philosophical position has a set of psychological mechanisms rooting for it (33).

In addition to diagnosing the resilience of the problems, experimental philosophy might also contribute to solving the problems, for we might discover that some of the natural responses are the product of error or bias. If so, we would have reason to disregard the responses rooted in error. On this topic, the current work is quite divided. One view is that people’s incompatibilist responses are the result of a deep misunderstanding of the nature of determinism (32). On this view, people mistakenly conclude that determinism entails that our psychological attitudes and values have no effect on decisions. As a result, people’s fear of determinism is ill founded. A very different account holds that people’s emotional reactions to reprehensible actions swamps their rational ability to recognize the threat that determinism poses for responsibility. When people are allowed to assess the problem of free will and responsibility in a calm fashion, they tend to judge that responsibility is indeed threatened by determinism (15). If this is the correct psychological explanation of the response, then people are right to think that determinism threatens moral responsibility.

Experimental philosophy is also poised to shed light on the belief in free will itself. If we identify why people think that their choices are not determined, we will be in a better position to evaluate that belief. A psychological account of the genesis of free-will beliefs probably cannot directly show that such beliefs are true or false. But knowing why people believe in free will might well put us in a position to evaluate whether or not people’s belief in free will is justified (34).

This Review has focused on the experimental philosophy of free will, but it is really a case study that gestures to the broader promise of experimental philosophy. As noted at the beginning of this paper, several other central philosophical

problems have their roots in everyday thought. People seem to be torn on whether morality has an objective footing or is merely relative to culture. What factors pull in those different directions? People also find it difficult to make sense of the idea that a physical object such as the brain could produce conscious experiences. Why is that connection challenging to understand? If we can comprehend the psychological sources of these kinds of problems, we might be in a better position to resolve or defuse the problems.

References and Notes

1. J. Knobe, S. Nichols, *Experimental Philosophy* (Oxford Univ. Press, New York, 2008).
2. T. Huxley, *Lessons in Elementary Physiology* (MacMillan, London, ed. 2, 1868).
3. T. Reid, *Essays on the Active Powers of Man* (G. G. J. and J. Robinson, London, 1788).
4. C. Campbell, *On Selfhood and Godhood* (George Allen and Unwin, London, 1957).
5. G. Strawson, *Freedom and Belief* (Oxford Univ. Press, Oxford, 1986).
6. R. Kane, *J. Philos.* **96**, 217 (1999).
7. P. van Inwagen, *An Essay on Free Will* (Oxford Univ. Press, Oxford, 1983).
8. G. Moore, *Ethics* (Oxford Univ. Press, Oxford, 1912).
9. S. J. Blakemore, D. M. Wolpert, C. D. Frith, *Trends Cogn. Sci.* **6**, 237 (2002).
10. D. M. Wegner, T. Wheatley, *Am. Psychol.* **54**, 480 (1999).
11. J. W. Moore, D. M. Wegner, P. Haggard, *Conscious. Cogn.* **18**, 1056 (2009).
12. L. E. Schulz, J. Sommerville, *Child Dev.* **77**, 427 (2006).
13. S. Nichols, *Mind Lang.* **19**, 473 (2004).
14. T. Kushnir, H. Wellman, N. Chernyak, in *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, N. Taatgen, H. van Rijn, Eds. (Cognitive Science Society, Austin, TX, 2009), pp. 87–92.
15. S. Nichols, J. Knobe, *Nous* **41**, 663 (2007).
16. H. Sarkissian et al., *Mind Lang.* **25**, 346 (2010).
17. B. Spinosa, *The Ethics* (Jan Rieuwertsz, Amsterdam, 1677) [S. Shirley, Transl. (Hackett, Cambridge, MA, 1992)].
18. R. Holton, *Philosopher’s Imprint* **6**, 1 (2006).
19. R. Nisbett, T. Wilson, *Psych. Rev.* **84**, 231 (1977).
20. J. Bargh, T. Chartrand, *Am. Psychol.* **54**, 462 (1999).
21. F. Nietzsche, *Beyond Good and Evil* (C. G. Naumann, Leipzig, Germany, 1886) [W. Kaufman, Transl. (Random House, New York, 1966)].
22. G. Strawson, *Philos. Stud.* **75**, 5 (1994).
23. D. Dennett, *Elbow Room* (MIT Press, Cambridge, MA, 1984).
24. D. Hume, *Treatise of Human Nature* (J. Noon, London, 1739; reprinted by Oxford Univ. Press, New York, 2000).
25. D. Pereboom, *Living without Free Will* (Cambridge Univ. Press, Cambridge, 2001).
26. F. De Brigard, E. Mandelbaum, D. Ripley, *Ethical Theory Moral Pract.* **12**, 511 (2009).
27. A. Roskies, S. Nichols, *J. Philos.* **105**, 371 (2008).
28. E. Nahmias, D. J. Coates, T. Kvaran, *Midwest Stud. Philos.* **31**, 214 (2007).
29. J. Goldberg, J. Lerner, P. Tetlock, *Eur. J. Soc. Psychol.* **29**, 781 (1999).
30. J. Lerner, J. Goldberg, P. Tetlock, *Pers. Soc. Psychol. Bull.* **24**, 563 (1998).
31. E. Nahmias, *J. Cogn. Cult.* **6**, 215 (2006).
32. E. Nahmias, D. Murray, in *New Waves in Philosophy of Action*, J. Aguilar, A. Buckareff, K. Frankish, Eds. (Palgrave-Macmillan, New York, 2011), chap. 9.
33. W. Sinnott-Armstrong, in *Experimental Philosophy*, J. Knobe, S. Nichols, Eds. (Oxford Univ. Press, New York, 2008), chap. 11.
34. S. Nichols, in *Are We Free: Psychology and Free Will*, J. Baer, J. Kaufman, R. Baumeister, Eds. (Oxford Univ. Press, Oxford, 2009), chap. 2.
35. I thank J. Knobe and T. Sommers for helpful comments.

10.1126/science.1192931