



Free will and responsibility

Eddy Nahmias*

Free will is a set of capacities for conscious choice and control of actions and is essential for moral responsibility. While determinism is traditionally discussed as the main potential challenge to free will and responsibility, other potential challenges exist and need to be considered by philosophers and scientists. The cognitive sciences are relevant to free will both to study how people understand free will and potential challenges to it, and to study whether these challenges are supported by relevant scientific evidence. © 2012 John Wiley & Sons, Ltd.

How to cite this article:

WIREs Cogn Sci 2012. doi: 10.1002/wcs.1181

INTRODUCTION

According to the traditional free will debate, research in the cognitive sciences is largely irrelevant to the question of whether humans have free will and are morally responsible for their actions. This view is mistaken. In this article, I diagnose this mistake (in the next section) and then correct it by discussing several ways that research in the cognitive sciences is highly relevant to debates about free will and responsibility.

First, research in psychology and the new field of experimental philosophy can help us systematize ordinary beliefs and intuitions about free will and responsibility, uncover psychological mechanisms that produce these intuitions, and examine how altering people's beliefs about free will alters their behavior (section *Beliefs About Free Will*). Such information is especially significant given the ever-increasing claims that research in neuroscience and psychology challenges the existence of free will. In the section *Potential Challenges to Free Will*, I consider whether such challenges from the cognitive sciences suggest the need to reassess the scope of human free will and the degree to which humans are morally responsible for their behavior.

THE TRADITIONAL FREE WILL DEBATE

Free will is usually and usefully defined as the set of capacities for choice and action control which are

essential for an agent to be morally responsible for his/her actions—for example, to deserve blame. The traditional free will debate has focused on one question: is free will compatible with *determinism*, the thesis that a description of the universe at one time, conjoined with the laws of nature, entails a description of the universe at any other time.¹ *Incompatibilists* argue that determinism entails that we cannot do otherwise, or that we cannot be the 'ultimate source' of our behavior, in ways required for free will. Incompatibilists then divide into *libertarians* who believe that we have free will (so determinism is false)^{2–5} and *hard determinists* who believe that, because determinism is true, we do *not* have free will.⁶ There are also *skeptics* about free will, who believe we do not have free will whether determinism or indeterminism is true,⁷ or who believe that free will is impossible.⁸ *Compatibilists*, on the other hand, argue that free will is possible even if determinism turns out to be true, either because determinism does not preclude the requisite ability to do otherwise or because neither such an ability nor 'ultimate sourcehood' is required for free will.^{9–13}

This debate turns largely on conceptual analysis of the relevant terms, such as 'ability' or 'sourcehood', and on the premises and principles of the respective arguments, but it is conspicuously distanced from empirical data—except to the extent that determinism is an empirical question. If any science is in a position to answer *that* question, it is physics, but not the sciences that study humans, since the truth of determinism depends on the fundamental laws of nature.¹⁴ If the laws of physics are indeterministic, then the cognitive sciences would not be in a position to establish the universal determinism required by incompatibilist

*Correspondence to: enahmias@gsu.edu

Department of Philosophy, Neuroscience Institute, Georgia State University, Atlanta, GA, USA

arguments. These sciences study systems that are not closed (e.g., human brains and behavior), so even if they are ‘near deterministic’ (a defective term), they are not cut off from all indeterministic interactions at lower levels (e.g., quantum events). For most incompatibilists, then, the question of whether *we* could have free will depends on an empirical question that might be answered by physicists, presumably while entirely ignoring humans. Even those libertarians who argue that free will requires indeterministic interactions in just the right processes in the brain still require the physics to turn out right.³

Incompatibilists argue that, *necessarily*, if determinism holds in a universe, *no* being has any free will, whether human, angel, animal, or alien. In practice, the discussion focuses on humans, but our specific cognitive and volitional capacities are often ignored. Meanwhile, a common compatibilist tactic is to describe the capacities of an ideal agent (e.g., fully rational and self-controlled) and argue that such capacities are both compatible with determinism and sufficient to secure free will.¹⁵ But the question of whether *actual* human beings have such idealized capacities—or the degree to which they have them—can become lost in the more abstract debate. Compatibilists typically suggest that establishing the compatibility of free will and determinism thereby establishes that humans have free will. This, of course, does not follow; establishing that agents *could* have free will in a deterministic universe says nothing about whether we humans *actually* have free will.

So, there are incompatibilists who believe we have free will and incompatibilists who believe we lack free will, and there are compatibilists who believe we have free will. There is a logical space to be filled: compatibilists who worry that we lack free will (see Figure 1). Regardless of whether free will is compatible with determinism, it may be incompatible with other possibilities. For instance, suppose *eliminativism* is true. That is the position that folk psychological concepts, such as desires and intentions—including those we employ when we deliberate and those we consider when we assess responsibility—should be eliminated, because they do not refer to anything; scientific discoveries will show that those concepts do not map onto the actual causes of human behavior.^{16,17} Or suppose *epiphenomenalism* is true. That is the position that conscious mental states play no causal role in behavior—for instance, because they are distinct from neural states that are causally sufficient for all behavior.^{18,19} Eliminativism and epiphenomenalism seem to challenge the existence of free will, since according to compatibilist and libertarian theories alike, free will requires that our conscious mental

states play an appropriate role in our behavior. As Robert Kane notes, ‘If conscious willing is illusory or epiphenomenalism is true, *all* accounts of free will go down, *compatibilist and incompatibilist*’.^{20,21}

These challenges are entirely independent from determinism. Epiphenomenalism and eliminativism do *not* depend on the truth of determinism, since they may be true in indeterministic universes. Conversely, the reality and causal efficacy of mental states is not ruled out by the truth of determinism; psychological states, both conscious and nonconscious, can exist and be causally efficacious even if their causal relations are subsumed under deterministic laws. As we will see below, some cognitive scientists muddy the waters by confusing determinism with these other potential challenges to free will.

Nonetheless, these scientists are right that empirical discoveries can challenge free will by challenging the causal role of consciousness or the degree to which humans possess the cognitive capacities essential to free and responsible agency. Most theories of free will suggest shared *necessary* conditions that include specific cognitive capacities. Indeed, while philosophers debating free will tend to highlight their differences, in fact they agree about many conditions required for free will. Most theories, whether compatibilist or libertarian, agree that free agents must have the cognitive capacities to consciously consider alternatives for action and to make choices based on their reasons for action. Conscious deliberation or choice need not accompany every free action, since agents may develop general principles for action (or character traits) that lead them to act without conscious reflection. And free agents need not always act on their best reasons. But most theorists agree that free will and responsibility require that agents possess the capacities to consciously consider their reasons and to govern their behavior in light of these reasons and have the opportunity to exercise those capacities at some relevant point before they act. Such conditions are shared by *compatibilists*, who emphasize, for instance, ‘the capacity for reflective self-evaluation’¹¹ and ‘reflective self-control’;²² *libertarians*, who say free action requires ‘a process of critical evaluation with respect to one’s conception of the good’⁵ and the capacities for ‘representing diverse, sophisticated plans of action’;⁴ and *skeptics* about free will, who say a free agent ‘knows the nature of [her] beliefs, desires and other mental states that bring about’ choices.²³ (see also Refs 3,8,10,12,13,15).

As these philosophers focus on their competing answers to the traditional debate about determinism, they neglect more fundamental potential challenges

		The possibility question: Is free will compatible with determinism?	
		YES (Compatibilism)	NO (Incompatibilism)
The actuality question: Do humans have free will?	YES	Compatibilists	Libertarians
	DEGREES		
	NO		Hard determinists and skeptics about free will

FIGURE 1 | The traditional questions and traditional positions in philosophical debates about free will. The unlabeled cells represent under-discussed but important positions, which are also the ones most likely to be informed by discoveries in cognitive science.

to their shared conditions for free will, challenges to human capacities for conscious deliberation and intention-formation and capacities for (moral) reasoning, and to our ability to exercise these capacities at a relevant point before action. Such challenges are especially relevant to questions about moral responsibility—for example, whether people *deserve* blame or punishment for bad actions or credit for good actions—since they apply to the capacities that allow people to recognize moral reasons for action and to control their behavior in response to such reasons.

The traditional free will positions offer ‘conclusive’ or all-or-nothing answers to the question of whether humans are free and responsible. But people possess cognitive capacities and are able to exercise them to varying degrees. So, if free will involves a set of capacities, then different individuals may possess more or less free will, and they may have greater or lesser opportunity to exercise free will in various situations. Information from cognitive science may suggest that we have *less* free will than we typically think without establishing that *none* of us has *any* free will (as incompatibilists argue determinism would show). This accords with our intuition that humans develop greater autonomy and control as they grow up, and it accords with our practices of holding people morally responsible to varying degrees, depending on the degree to which they have matured to possess the relevant cognitive and volitional capacities, and on the degree to which people have opportunities to exercise those capacities in action (e.g., we tend to mitigate responsibility when people are under undue cognitive or emotional stress).

So, even if the cognitive sciences are not in a position to contribute to the traditional debate about whether free will is compatible with determinism or to answer the question of whether universal determinism is true, these sciences are well situated to inform us about cognitive capacities essential to free and responsible agency and about the degree to which humans possess them (see Figure 1).

As I will discuss below, some scientists suggest that empirical research shows that free will is an illusion, but this conclusion depends on how we understand free will. Indeed, according to some conceptions of free will, research in cognitive science can help to explain how free will works, rather than explaining *away* free will. So, it is also essential to study how people actually understand free will and moral responsibility, as well as how their beliefs and behaviors might alter in response to scientific claims about free will.

BELIEFS ABOUT FREE WILL

Philosophers sometimes appeal to claims about what is intuitive to ordinary people to support their theories of free will, and scientists who discuss free will typically go further and simply define free will in terms of their assumptions about how people understand it. While it is a mistake to define free will solely in terms of folk usage or intuitions, ‘free will’ is also not a technical or scientific concept whose definition can be stipulated by experts without considering the way people actually use the term and the roles it plays in people’s attributions of responsibility.

Empirical information about ordinary beliefs can contribute to debates about free will in at least four ways: (1) *Other things being equal*, an intuitive view should not be replaced by a counterintuitive view without good reason,^{24,25} and folk intuitions should serve as input to a process of *reflective equilibrium* that seeks a stable position among such intuitions, information from relevant sciences, and relevant normative concerns.^{26,27} (2) Studies that systematically alter features of cases to examine ensuing effects on people’s judgments about free will and responsibility provide information about the psychological mechanisms generating such judgments and about why people may generate mistaken judgments.^{28,29} (3) To determine if certain discoveries challenge free will, we must determine whether people

in fact take free will to be incompatible with such discoveries (see section *Potential Challenges to Free Will*). (4) Cognitive science can provide information not only about what might threaten free will but also about how people will react to such news—that is, how their beliefs and behavior change in response to scientific claims about free will.

Having motivated the importance of empirical information about people's beliefs about free will, I will briefly survey what the research has uncovered so far. But this research is less than a decade old, so it is preliminary and exploratory. Most of it has been carried out by 'experimental philosophers', sometimes in collaboration with psychologists and other cognitive scientists. The field is ripe for new interdisciplinary research.

Folk Intuitions About Incompatibilism

Incompatibilists often suggest that their view is intuitive and that compatibilism is, as William James suggested, a 'quagmire of evasion'. Robert Kane writes: 'In my experience, most ordinary persons start out as natural incompatibilists'³⁰ (p. 217). And Galen Strawson contends that the incompatibilist conception of free will (though impossible to satisfy) 'is just the kind of freedom that most people ordinarily and unreflectively suppose themselves to possess'⁸ (p. 30). Initial work in experimental philosophy on free will sought to test these claims.

To examine laypersons' intuitions about the relationships between determinism and free will and moral responsibility, Nahmias et al. developed various scenarios that describe determinism and asked people whether agents in such scenarios can be free and responsible.^{24,31} For instance, one study presented participants with this scenario based on Laplace's definition of determinism in terms of perfect prediction:

Imagine that in the next century we discover all the laws of nature, and we build a supercomputer which can deduce from these laws of nature and from the current state of everything in the world exactly what will be happening in the world at any future time. It can look at everything about the way the world is and predict everything about how it will be with 100% accuracy. Suppose that such a supercomputer existed, and it looks at the state of the universe at a certain time on March 25th, 2150 A.D., twenty years before Jeremy Hall is born. The computer then deduces from this information and the laws of nature that Jeremy will definitely rob Fidelity Bank at 6:00 PM on January 26th, 2195. As always, the supercomputer's prediction is correct; Jeremy robs Fidelity Bank at 6:00 PM on January 26th, 2195.

A significant majority of participants (76%) judged that Jeremy has free will and that he is morally responsible (83%). Similar results were found in variations with praiseworthy and morally neutral actions and in two other scenarios using different descriptions of determinism. These results put pressure on the assumption that incompatibilism is the intuitive position.

Later studies complicate the picture. Nichols and Knobe found that when people read their description of a fully deterministic universe (*A*) and a universe in which human decisions are not fully determined (*B*), most responded that our universe is like universe *B* and also that individuals in universe *A* could *not* be 'fully morally responsible' for their behavior. But when the scenario included a description of a person in universe *A* who kills his family, most people responded that he *is* fully morally responsible for his behavior. Nichols and Knobe argue that these results suggest that people have an incompatibilist *theory* of responsibility which they then apply incorrectly in concrete, emotionally laden cases.²⁸ Their research initiated investigations into the psychological *sources* of people's conflicting intuitions.

Psychological Sources of Intuitional Conflict

Regardless of whether majorities of nonphilosophers express intuitions consistent with specific philosophical theories, studies carried out so far indicate that (1) different people make different judgments about the same scenarios; (2) minor variations to scenarios can produce significant differences in response patterns; and (3) in some cases, the same individual will respond in ways that suggest he/she has conflicting intuitions about the issues. That is, there appear to be interesting *inter*-subject and *intra*-subject intuitional conflicts about the relationships between free will, moral responsibility, and determinism. Careful psychological investigation, informed by philosophers' conceptual resources, is needed to elucidate the contours of these conflicts and perhaps uncover the psychological mechanisms driving them. Such work could have important implications for the philosophical debates; for instance, it may challenge the assumption shared by most philosophers that there are invariant conditions for proper judgments of moral responsibility,³² and it could establish in what ways a theory of freedom or responsibility *revises* our folk theory and practices and in what ways it *preserves* them.³³

So far, experimental philosophers have discovered several interesting patterns of responses. As mentioned, Nichols and Knobe have discovered that emotional salience increases judgments of

agents' responsibility, including agents in deterministic scenarios. Second, they and others have found that concrete cases—ones with specific agents performing specific actions—are more likely to evoke judgments of responsibility and free will than abstract cases—ones with only general descriptions of determinism.^{28,34,35} One might argue that judgments in abstract cases are more reliable and hence provide better information about folk theories of responsibility. Or one might argue that concrete cases (at least ones that do not overly arouse emotion) are more reliable, because they engage 'theory of mind' capacities or other psychological processes plausibly relevant for accurately assessing agency and responsibility. Third, Nahmias and Murray have found that judgments of free will and responsibility depend on how one interprets a particular description of determinism. If a person interprets determinism to entail 'bypassing' of one's mental states—for example, that agents' beliefs, desires, or decisions play no causal role in what they do—then she is likely to judge that agents lack free will and responsibility. If a person does not interpret determinism to entail such bypassing, she is likely to judge that agents can have free will and responsibility in deterministic scenarios. Since determinism does *not* entail bypassing or epiphenomenalism about mental states, Nahmias and Murray argue that these results provide an error theory for folk intuitions that *appear* to support incompatibilism: when people properly understand determinism, they take it to be compatible with free will and responsibility.^{29,36}

One source of these bypassing judgments likely derives from our ignorance about the mind-body relation. Neither cognitive scientists nor philosophers have developed an established or comprehensible theory of how mental states or processes, including conscious ones, are identical to—or can be explained in terms of—physical (e.g., neural) states or mechanistic processes. So, when decisions or actions are described as completely caused by physical processes and laws, people readily interpret this to mean that our mental states are bypassed. If the brain does it all, and we don't understand how the mind is the brain, then it seems the mind has nothing left to do. If this interpretation of the psychological sources of our intuitions about free will is plausible, then it helps to explain why both scientists and laypersons readily interpret neuroscientific explanations of human decision-making as threatening free will (see section *Potential Challenges to Free Will*). This interpretation also predicts that, if and when we gain a better neuroscientific understanding of conscious mental states, their relation to brain states, and their causal role in behavior, people will be less likely to find

neuroscientific explanations threatening to human agency and responsibility and more likely to see them as explanations for how free will works.

Unless and until such explanations are forthcoming, however, telling people that science can completely explain human behavior in reductive or mechanistic terms is likely to diminish their belief in free will. Psychologists are now exploring how such changes in belief can also influence people's behavior.

Behavioral Effects of Altering Beliefs About Free Will

Crick once wrote: "You", your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behaviour of a vast assembly of nerve cells and their associated molecules'.³⁷ Psychologists Kathleen Vohs and Jonathan Schooler ran an experiment in which they gave an extended version of Crick's discussion of free will to one group of participants, while a control group received a Crick passage that does not discuss free will. Participants then took a math test set up so that they could cheat to see the right answers. Those who received the no-free-will prime cheated significantly more than controls (about 14 times out of 20 questions compared to 10). Vohs and Schooler also found that diminishing belief in free will led participants to lie about how many problems they got right to receive larger rewards.³⁸ Baumeister et al. have found that diminishing belief in free will led participants to help others less and to be more aggressive toward others.³⁹ Further studies suggest that reduced belief in free will correlates with worse job performance, less counterfactual thinking, even diminished brain activity associated with voluntary movement, while it also leads people to be more forgiving and less punitive.⁴⁰

These interesting, but preliminary, findings suggest that people's beliefs about free will can have important effects on their behavior, but important questions remain about what specific information drives the changes in belief, what cognitive processes drive the behavioral changes (e.g., diminished belief in responsibility, diminished willpower), and how long-lasting such effects might be. Since free will is typically taken as a condition for moral responsibility and desert, so reducing people's belief in free will might also lead to changes in the way we treat wrongdoers. For instance, some philosophers and scientists believe that humans lack libertarian free will and that such free will is required to deserve retributive blame or punishment.^{7,41} They argue that a scientifically

informed understanding of human decision-making should lead us to give up our legal system's retributive justification for punishment, instead focusing entirely on 'forward-looking' considerations, such as rehabilitation and deterrence, which need not be justified in terms of criminals' free will. Others argue that neither our legal system nor our practices of retributive punishment require any conception of free will that is challenged by determinism or by existing scientific evidence about human decision-making.⁴² Resolving these debates will require both philosophical and legal arguments but also more empirical information about people's understanding of free will, blame, and punishment.

We have seen that people's beliefs about free will likely influence their behavior and may impact assessments of responsibility, including blame and legal punishment. The existing evidence suggests that beliefs about free will are unlikely to be influenced by proclamations from philosophers or physicists about *determinism*; rather, they are more likely to be influenced by claims that neuroscience or psychology establishes that free will is an illusion. So, we need to consider whether or not these sciences in fact provide evidence that challenges free will and responsibility.

POTENTIAL CHALLENGES TO FREE WILL FROM COGNITIVE SCIENCE

Increasingly, cognitive scientists are suggesting that their research shows that free will is an illusion. Psychologist Bargh writes, 'The phenomenological feeling of free will is very real... but this strong feeling is an illusion'⁴³ (pp. 148–149). Neuroscientist Haynes writes: 'Our decisions are predetermined unconsciously a long time before our consciousness kicks in... It seems that the brain is making the decision before the person themselves... This doesn't rule out free will, but it does make it implausible'.⁴⁴ Cognitive scientists Greene and Cohen conclude, 'The net effect of this influx of scientific information will be a rejection of free will as it is ordinarily conceived with important ramifications for the law'⁴¹ (p. 1776). To determine whether discoveries in the cognitive sciences in fact challenge human free will and responsibility requires, first, establishing criteria for free will and responsibility, and second, examining whether these discoveries challenge those particular criteria.

As we have seen, according to incompatibilist theories, one criterion for free will and responsibility is the falsity of determinism. Some scientists assume that incompatibilism is obviously true and is entailed by ordinary beliefs about free will, and then they argue that cognitive science is discovering that determinism

is true, or at least true for all human behavior. But, as explained above, it is debatable whether ordinary intuitions take determinism, properly construed, to be incompatible with free will, and most philosophers are compatibilists who argue that it is not. Furthermore, even if determinism were incompatible with free will, cognitive science is not in a position to establish the truth of determinism. As it turns out, when cognitive scientists argue that their discoveries challenge free will, even if they use the language of determinism, they typically have something else in mind.

The Role of Consciousness in Action

We have seen that most theories of free will, compatibilist and incompatibilist alike, take it to require that consciousness plays the right role in action. One way that cognitive scientists have suggested their discoveries might challenge free will is by challenging the causal role of conscious mental processes in action. Again, this threat to free will is consistent with both the truth or falsity of determinism.

This challenge was raised initially by Benjamin Libet's experiments.⁴⁵ Libet found that voluntary muscle movements (e.g., flexing one's wrist) are preceded by a 'readiness potential' (RP), a brain wave that occurs about half a second (500 ms) before the movement. But Libet's subjects reported being aware of the 'intention, desire, or urge' to move only about 150 ms before the movement—350 ms *after* the RP. Libet concludes that voluntary actions 'begin in the brain unconsciously, well before the person consciously knows he wants to act'⁴⁶ (p. 51), and he interprets this result to show that our conscious intention to move is *not* the cause of our movement but, like the movement itself, an effect of earlier (non-conscious) brain activity. Haynes et al. have extended Libet's findings. Using fMRI, they discerned patterns of neural activity that correlated with subjects' decision to press either a right or left button up to 7–10 s before they were aware of making their decisions. They take these findings to bolster Libet's case that the 'subjective experience of freedom is no more than an illusion and that our actions are initiated by unconscious mental processes long before we become aware of our intention to act'⁴⁷ (p. 543). These discoveries appear to reduce the role of consciousness to *observing* our decisions rather than *making* them.

Daniel Wegner extends this model to suggest that conscious will is an illusion.⁴⁸ While we think that our experience of consciously willing our actions is indicative of how our actions are caused, Wegner argues that we are systematically mistaken, that the evidence shows that 'the real causes of human

action are unconscious⁴⁸ (p. 97). Our experience of conscious will results from having relevant conscious thoughts (e.g., intentions) just prior to the action, while being unaware of any competing causes of the action, but following Libet, Wegner argues that the thoughts are themselves caused by prior (nonconscious) brain activity such that a conscious intention 'might just be a loose end—one of those things, like the action, that is caused by prior brain and mental events'⁴⁸ (p. 55). The evidence for this model of 'apparent mental causation' is based primarily on cases where people lack an experience of consciously willing a bodily movement that they in fact brought about (e.g., automatisms, hypnosis, alien hand) and experiments where people experience some sense of agency for a movement or an outcome that they do not in fact cause. These seemingly exceptional cases Wegner takes to represent the rule: our conscious intentions *never* cause our actions.

There have been numerous responses to Libet's and Wegner's empirical evidence and the implications they draw from it.^{49–53} One problem is that the scientists interpreting these experiments often implicitly or explicitly assume a dualist view that conscious mental states are nonphysical and only influence action by influencing physical processes (such as neural activity) that precedes behavior. Libet writes, 'But we have not answered the question of whether our consciously willed acts are fully determined by natural laws that govern the activities of nerve cells in the brain, or whether conscious decisions can proceed to some degree independently of natural determinism [sic]'⁴⁵ (p. 55). Wegner suggests an answer to Libet's question: 'the brain started first, followed by the experience of conscious will, and finally followed by action'⁴⁸ (p. 55). However, there is no reason to begin with such dualistic assumptions. Our experiences of voluntary action do not tell us that conscious processes are distinct from brain processes. Our experiences are 'topic-neutral' among competing metaphysical theories; they do not commit us to dualism (nor rule it out), and they are consistent with the theory that consciousness is realized in (or identical to) certain brain states. Assuming that mental states are realized in brain states—an assumption at the core of cognitive science—then the question of whether conscious processes play a causal role in action turns on whether those brain states that realize consciousness play the right sort of causal role in our actions, not whether our conscious mental states are regularly preceded by any brain states.

With that question in place, one might still argue that the data from research like Libet's and Haynes' show that those brain states that realize consciousness

are *not* causes of our actions, since they occur too late to influence action. However, this conclusion is too hasty. These data are consistent with other interpretations. For instance, the early brain activity measured in these experiments (such as RPs) may represent non-conscious *urges* to flex soon, rather than fully formed *intentions* or *decisions* to flex. If so, then the activity might be among the causes of conscious intentions (and their underlying neural correlates), which are experienced a few hundred milliseconds later, while sometimes the urge is 'vetoed,' perhaps by a conscious intention not to act on that particular urge (in Haynes' study, the early brain activity only predicted which button participants press at 10% above chance). This interpretation, if true, allows that conscious proximal intentions can still causally influence when and how the person acts.^a Another possibility is that early nonconscious brain activity *is* the correlate of the conscious intention, but *reporting* the intention involves separate conscious processes that take additional time. Or the early brain activity may represent preparatory activity for consciously experienced intentions or decisions; if we assume that conscious processes are realized in brain processes, we should expect that conscious experiences do not arise out of nowhere and in no time. For instance, in these experiments participants presumably consciously processed the experimenters' instructions, which in Haynes' study were 'to press either the left or right button with the index finger of the corresponding hand immediately when they became aware of the *urge* to do so' and 'to avoid any form of preplanning for choice of movement or time of execution' (Libet's instructions were similar). If participants followed these instructions, they formed a distal intention (or plan) to allow urges to press one of the buttons to arise within them and pay attention to when they arise. This (conscious) distal intention presumably causally influenced the spontaneous generation of nonconscious urges to act in one of the two allowed ways. Furthermore, subjects may be reporting not a consciously formed *intention* to act now, but rather the time at which they became aware of the urge to act. Given that these experiments involve dozens of trials, it is even more plausible that subjects develop an action plan to allow urges to come upon them and (typically) to let those urges proceed to action, while also carrying out the (nonnormal) task of attending to when each urge occurred.

In support of these interpretations, Pockett and Purdy found differences in participants' reports of the time of awareness, depending on whether they were asked to report when they experience an *urge* to press one of two buttons or when they made a *decision* to press one of the buttons. The event-related potentials

(ERPs) for the different trials were also different. Furthermore, many participants reported awareness of decisions well before awareness of urges and before Libet's RP onset of about 500 ms before movement.⁵⁴ Trevena and Miller also present results suggesting that the RP is not a correlate of a *decision* to move but of *preparation* for a decision either to move or not to move.⁵⁵

Wegner, for his part, presents no relevant neuroscientific data (other than Libet's) to suggest that the brain processes associated with our conscious intentions are causally cut off from those that produce actions. The exceptional cases from the psychological literature that he cites show only that the experience of will is not always veridical, not that it is *never* veridical. Without the neuro-anatomical data to demonstrate that the relevant brain processes are causally unconnected, the best interpretation for these 'illusions of will' is by analogy with visual illusions, which certainly do not show that our visual experiences are systematically mistaken. Indeed, as with most visual illusions, explanations for illusions of will may be given in terms of a generally reliable system sometimes producing inaccurate output because of some unusual feature of the situation. The fact that we *sometimes* perform complex behaviors *without* conscious intentions (e.g., under hypnosis) does not show that on the many occasions we perform complex behaviors *with* conscious intentions, those intentions are causally irrelevant.⁵²

Nonetheless, the relevant evidence might come in to show that when we consciously intend an action immediately before we act, our being conscious (and the relevant neural correlates) simply occurs too late (or the neural correlates occur 'in the wrong place') to causally influence the action. Even *if* this turns out to be true, however, it may not represent a significant threat to free will. The common feature of the theories of free will I outlined above is not that conscious intentions *just prior* to action (or their neural correlates) causally influence our actions. Rather, what is more important is that conscious deliberations, plans, and distal intentions (or their neural correlates) can have a proper downstream effect on how we act in the relevant situations. If such conscious causation is in place, it would allow that we can act in accord with reasons that we have (at some point) consciously considered and accepted. There is no neuroscientific evidence (yet) to show that neural processes involved in conscious deliberation, reasoning, and planning have no such effects on what we do or that our conscious monitoring of our behavior in light of such reasons or plans is not critically involved in how we carry out and adjust our actions. On the contrary, some evidence from

psychology suggests that conscious planning, reasoning, and intention-formation does influence actions. For instance, Peter Gollwitzer's work on implementation intentions shows that people are more likely to follow through on a plan when they consciously form an intention to act at a certain time than when they do not form such an intention.⁵⁶ And Baumeister et al.'s work provides several examples of behaviors that are improved by conscious reasoning and conscious attention to action.⁵⁷

It remains an open question, however, whether there are significant limitations to the role of conscious mental processes in action and to our capacities for rational control of action. For instance, research in social psychology suggests that we are often influenced by situational factors that we do not recognize and that we typically reject as reasonable influences on us, such as the strong effect of inactive bystanders on diminishing helping behavior. There is a long list of such situational factors and cognitive biases that influence behavior, from seemingly irrelevant factors, such as ambient noises or smells, to factors most of us wish were less influential, such as blind obedience to authority, as demonstrated by Milgram's infamous experiments.^{58,59} Such research also suggests that character traits are not robust or stable across various situations, nor are traditional character traits good predictors of behavior,⁶⁰ and that our explanations of our own and others' actions are often based on mistaken folk theories and inaccurate introspection.^{61,62} Hence, our capacities to regulate our behavior in accord with traits we have inculcated or reasons we consciously adopt may be limited. Understanding whether and when our decisions and actions are driven by, or even consistent with, reasons and desires we have consciously endorsed has important implications for understanding the scope of human free will and responsibility.⁶³ Research in cognitive science will help to illuminate these issues.

CONCLUSION

The traditional free will debate has focused on whether free will is compatible with determinism, a debate that shows little sign of resolution and that is unlikely to be informed by discoveries in cognitive science. Whether our decisions are deterministically caused, however, is not the only significant question regarding free will and moral responsibility. It is also important to consider the specific causal explanations for our actions, and here the cognitive sciences can play a central role. If free and responsible agency depends on a set of cognitive and volitional capacities, possessed and exercised to varying degrees, then the cognitive

sciences can inform us about the extent to which humans (as a species and as individuals) possess those capacities and are able to exercise them in various situations. Delineating the contours of our freedom will require both philosophical analysis and scientific investigation.

Another type of investigation at the crossroads of philosophy and cognitive science explores what people believe about free will and moral responsibility so that we can better understand ordinary intuitions about these issues, explore what drives people's conflicting intuitions, and study how people's beliefs about their own and others' freedom influence their behavior and their attributions of moral responsibility. This last issue is especially relevant, since scientific discoveries about human agency have the potential to challenge the existence, or at least the scope, of free will, and these discoveries are being publicized widely. It is crucial that we understand which discoveries properly challenge free will, the ways they may impact our beliefs about free will and responsibility, and the ways they may thus influence our behavior and our society's moral and legal practices.

Finally, while research in cognitive science is often presented as challenging free will, it also has the

potential to *explain* how free will works rather than explaining it *away*—to explain how conscious deliberation and planning affects our choices, how we can be responsive to reasons, and how decision-making and intentional action work. Thus, several strands of research in cognitive science that have not been highlighted in this article also inform the nature and scope of human freedom, including research on the cognitive and neurobiological processes involved in voluntary action, deliberation, and planning,^{64,65} as well as our experiences of freedom, agency, and control.^{51,66,67}

NOTES

^aThis alternative account is consistent with Libet's own view that the conscious will has 'veto power,' but the way he describes this possibility seems to require dualism.⁴⁶ Note that my discussion has ignored arguments against the causal efficacy of mental states that supervene on physical states (e.g., Jaegwon Kim's causal exclusion argument⁶⁸), but the soundness of these metaphysical arguments is distinct from neuropsychological data such as Libet's. The work of Libet and Wegner does not bear directly on these philosophical debates about mental causation.

REFERENCES

- van Inwagen P. *An Essay on Free Will*. Oxford: Clarendon Press; 1983.
- Chisholm R. *Human Freedom and the Self*. Lawrence, KA: University of Kansas; 1964.
- Kane R. *The Significance of Free Will*. New York: Oxford University Press; 1996.
- O'Connor T. *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press; 2000.
- Ekstrom L. *Free Will: A Philosophical Study*. Boulder: Westview Press; 2000.
- Baron d'Holbach PH. *System of Nature; or, The Laws of the Moral and Physical World*. London: G. Kearsley; 1797.
- Pereboom D. *Living Without Free Will*. Cambridge: Cambridge University Press; 2001.
- Strawson G. *Freedom and Belief*. Oxford: Clarendon Press; 1986.
- Campbell JK. A compatibilist theory of alternative possibilities. *Philos Stud* 1997, 88:319–330.
- Fischer JM, Ravizza M. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press; 1998.
- Frankfurt H. *The Importance of What We Care About*. Cambridge: Cambridge University Press; 1988.
- Wolf S. *Freedom Within Reason*. New York: Oxford University Press; 1990.
- Dennett D. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: MIT Press; 1984.
- Hofer C. Causal determinism. In: Zalta EN, ed. *The Stanford Encyclopedia of Philosophy*. Spring; 2010. Available at: <http://plato.stanford.edu/archives/spr2010/entries/determinism-causal/>. (Accessed March 25, 2012).
- Mele A. *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press; 1995.
- Churchland P. Eliminative materialism and propositional attitudes. *J Phil* 1981, 78:67–90.
- Stich S. *Deconstructing the Mind*. New York: Oxford University Press; 1996.
- Robinson WS. Epiphenomenalism. *WIREs Cogn Sci* 2010, 1:539–547.
- Huxley TH. *On the Hypothesis that Animals are Automata, and its History*. *The Fortnightly Review* 1874, 16 (New Series): 555–580. Reprinted in *Method and Results: Essays by Thomas H. Huxley*. New York: D. Appleton and Company; 1898.

20. Kane R. *Remarks on the Psychology of Free Will*. Winston-Salem, NC: The Society Of Philosophy And Psychology, Wake Forest University; 2005.
21. Berofsky B. *Autonomy Without Free Will*. Cambridge: Cambridge University Press; 2005, 82.
22. Wallace J. *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press; 1993.
23. Double R. *The Non-Reality of Free Will*. New York: Oxford University Press; 1991.
24. Nahmias E, Morris S, Nadelhoffer T, Turner J. Is Incompatibilism Intuitive? *Phil Phenomenol Res* 2006, 73:28–53.
25. Lycan W. Free will and the burden of proof. In: O'Hear A, ed. *Proceedings of the Royal Institute of Philosophy for 2001-02*. Cambridge: Cambridge University Press; 2003, 107–122.
26. Swanton C. *Freedom: A Coherence Account*. Indianapolis: Hackett Publishing; 1992.
27. Murray D, Nahmias E. Explaining away incompatibilist intuitions. Unpublished manuscript.
28. Nichols S, Knobe J. Moral responsibility and determinism: the cognitive science of folk intuitions. *Noûs* 2007, 41:663–685.
29. Nahmias E, Murray D. Experimental philosophy on free will: an error theory for incompatibilist intuitions. In *New Waves in Philosophy of Action*. New York: Palgrave-Macmillan; 2010, 189–215.
30. Kane R. Responsibility, luck, and chance: reflections on free will and indeterminism. *J Phil* 1999, 96:217–240.
31. Nahmias E, Morris S, Nadelhoffer T, Turner J. Surveying freedom: folk intuitions about free will and moral responsibility. *Phil Psychol* 2005, 18:561–584.
32. Knobe J, Doris J. *Moral Responsibility*. Oxford: Oxford University Press; 2011.
33. Vargas M. The revisionist's guide to moral responsibility. *Phil Stud* 2005, 125:399–429.
34. Nichols S, Roskies A. Bringing moral responsibility down to earth. *J Phil* 2008, 105:371–388.
35. Nahmias E, Coates DJ, Kvaran T. Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Stud Phil* 2007, 31:214–242.
36. Nahmias E. Intuitions about Free Will, Determinism, and Bypassing. In: *The Oxford Handbook on Free Will*. 2nd edn. Oxford: Oxford University Press; 2011, 555–575.
37. Crick F. *The Astonishing Hypothesis*. New York: Scribner's; 1994.
38. Vohs K, Schooler J. The value of believing in free will: encouraging a belief in determinism increases cheating behavior. *Psychol Sci* 2008, 19:49–54.
39. Baumeister RF, Masicampo EJ, DeWall CN. Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personal Social Psychol Bull* 2009, 35:260–268.
40. Rigoni D, Kühn S, Sartori G, Brass M. Inducing disbelief in free will alters brain correlates of preconscious motor preparation: the brain minds whether we believe in free will or not. *Psychol Sci* 2011, 22:613–618.
41. Greene JD, Cohen JD. For the law, neuroscience changes nothing and everything. *Phil Trans R Soc Lond B (Special Issue on Law and the Brain)* 2004, 359:1775–1785.
42. Morse SJ. Determinism and the death of folk psychology: two challenges to responsibility from neuroscience. *Minnesota J Law, Sci Technol* 2008, 9:1–35.
43. Bargh J. Free Will is Un-natural. In: *Are We Free? Psychology and Free Will*. New York: Oxford University Press; 2008, 128–154.
44. Hotz RL. Get Out of Your Own Way. New York: Wall Street Journal; 2008. Available at: <http://online.wsj.com/article/SB121450609076407973.html>. (Accessed March 26, 2012).
45. Libet B. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav Brain Sci* 1985, 8:529–566.
46. Libet B. *Do We Have Free Will?* Exeter: Imprint Academic; 1999.
47. Soon C, Brass M, Heinze H, Haynes J. Unconscious determinants of free decisions in the human brain. *Nature Neurosci* 2008, 11:543–545.
48. Wegner D. *The Illusion of Conscious Will*. Cambridge: MIT Press; 2002.
49. Mele A. *Effective Intentions: The Power of Conscious Will*. New York: Oxford University Press; 2009.
50. Nahmias E. When consciousness matters: a critical review of Daniel Wegner's *The Illusion of Conscious Will*. *Phil Psychol* 2002, 15:527–541.
51. Bayne T. Phenomenology and the Feeling of Doing: Wegner on the Conscious Will. In: *Does Consciousness Cause Behavior? An Investigation of the Nature of Volition*. Cambridge: MIT Press; 2006.
52. Nahmias E. Agency, authorship, and illusion. *Conscious Cogn* 2005, 14:771–785.
53. Sinnott-Armstrong W, Nadel J, eds. *Conscious Will and Responsibility*. New York: Oxford University Press; 2011.
54. Pockett S, Purdy S. Are voluntary movements initiated preconsciously? The relationships between readiness potentials, urges, and decisions. In *Conscious Will and Responsibility*. New York: Oxford University Press; 2011, 34–46.
55. Trevena J, Miller J. Cortical movement preparation before and after a conscious decision to move. *Conscious Cogn* 2002, 11:162–190.
56. Gollwitzer PM. Implementation intentions. *Am Psychol* 1999, 54: 493–503.
57. Baumeister RF, Masicampo EJ, Vohs KD. Do conscious thoughts cause behavior? *Annu Rev Psychol* 2011, 62:331–361.

58. Ross L, Nisbett R. *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill; 1991.
59. Milgram S. *Obedience to Authority*. New York: Harper and Row Publishers; 1969.
60. Doris J. *Lack of Character*. Cambridge: Cambridge University Press; 2002.
61. Haidt J. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol Rev* 2001, 108:814–834.
62. Nisbett R, Wilson T. Telling more than we can know: verbal reports on mental processes. *Psychol Rev* 1977, 84:231–259.
63. Nahmias E. Autonomous Agency and Social Psychology. In: *Cartographies of the Mind: Philosophy and Psychology in Intersection*. Dordrecht: Springer; 2007, 169–185.
64. Pockett S, Banks WP, Gallagher S, eds. *Does Consciousness Cause Behavior?* Cambridge: MIT Press; 2006.
65. Baumeister R, Mele A, Vohs K. *Free Will and Consciousness: How Might They Work*. New York: Oxford University Press; 2010.
66. Horgan T, Tienson J, Graham G. The Phenomenology of First-person Agency. In: *Physicalism and Mental Causation*. Exeter: Imprint Academic; 2003.
67. Nahmias E, Morris S, Nadelhoffer T, Turner J. The phenomenology of free will. *J Conscious Stud* 2004, 11:162–179.
68. Kim J. *Mind in a Physical World*. Cambridge: Bradford; 2000.

FURTHER READING

- Baer J, Kaufmann J, Baumeister R, eds. *Are We Free? Psychology and Free Will*. New York: Oxford University Press; 2008.
- Dennett D. *Freedom Evolves*. New York: Viking; 2003.
- Kane R, ed. *The Oxford Handbook on Free Will*. 2nd ed. New York: Oxford University Press; 2011.
- Swinburne R, ed. *Free Will and Modern Science*. New York: Oxford University Press; 2012.
- Watson G, ed. *Free Will*. 2nd ed. New York: Oxford University Press; 2003.