

Can Neuroscience Inform the Free Will Debate?

Brandi Jo Newell

Psychology

Wellesley College

Armchair philosophers, professional philosophers and even reluctant, first-year students in introductory philosophy classes have all been drawn in by the tremendously interesting and equally frightening question: *Do we have free will?* Some find solace in Cartesian dualism,¹ but the rest of us monists² are left to grapple with the inconsistency between our intuitive feeling that we are free to make our own decisions and the knowledge that our minds operate within what science assumes to be a deterministic universe. For millennia, this question has been approached theoretically, but new neuroscientific findings may give researchers the ability to broach the subject experimentally. The discussion has not just moved from the salon to the laboratory, but also to the newsroom, where lay press articles have been cropping up with ominous titles like “Free Will: Now You Have It, Now You Don’t” (Overbye, 2007), “The Brain on the Stand” (Rosen, 2007), and “Case Closed on Free Will?” (Youngsteadt, 2008). Undoubtedly, the revived academic and public interest in this age-old debate stems from the profound moral and legal implications that rest on the idea that humans have free will.

This paper seeks to establish whether or not the new findings can, in fact, inform the free will debate and, if so, what that means for society’s conceptions of moral and legal responsibility. The first section draws the battle lines by explaining both sides of the debate and defining the terms that will be used for the remainder of the paper. The second section takes up the question of neuroscience’s ability to weigh in on the dispute, and the third addresses the possible moral and legal ramifications.

Section 1: Free will, Determinism and Compatibilism

The standard free will debate has been framed as incompatibilist—either the world is deterministic, precluding free will (as so-called “hard determinists” believe), or it is not, and

¹ Rene Descartes proposed that the mind and body are made of separate but interacting substances. Because the mind is a non-physical substance, it is not subject to the physical laws that govern the rest of the world (and, therefore, its actions are not governed by deterministic principles). Due to this separation, dualists can believe in both determinism and free will, escaping the debate all together.

² Monists believe that the world is comprised of only one substance—a physical substance, in the case of materialists—as opposed to separate substances for mind and body.

free will is possible (as “libertarians”³ would argue). In other words, free will and determinism have been pitted against each other as mutually exclusive opposites. Others (aptly referred to as “compatibilists”) disagree and see free will and determinism as entirely compatible.

I argue that free will and determinism are tightly interwoven concepts, albeit not necessarily opposites. Pulling their meanings apart from one another can be helpful in deciding if and how neuroscientific research is relevant to the free will debate. Further, I contend that modern compatibilist arguments are flawed, and that free will (as we intuitively think of it) cannot exist in a determined world.

Determinism

Determinism has been defined as the idea that every event is caused by an antecedent event and governed by the laws of nature (Van Inwagen, 1982). The opposite of determinism, then, would be a belief that every event is *uncaused* and is *not* governed by the laws of nature. In other words, chaos or randomness (not free will!) is the opposite of determinism. Determinism is best conceptualized as a spectrum with determinism on one side, chaos on the other, and different degrees to which events follow laws and causal relationships in between.

While it is very uncomfortable to think that all of our decisions are determined by events set into motion millions of years ago and that it is not really the “self” who is in charge, it may be even more uncomfortable to think of our decisions as *undetermined*. As many philosophers have pointed out (Churchland, 2004; Dennett, 1984; Hume, 1739; Wegner, 2002;), no one would be happy if her decisions were entirely random—unrelated to her beliefs, desires and temperament; we would not think of these decisions as being “free,” any more than we would events that were entirely determined by physical laws. Note that this is not good news for the libertarian argument, which holds that free will can exist because determinism is false.

In addition to the fact that indeterminism is no more satisfying than determinism for either side of the argument, it also lacks any form of evidentiary support. There is ample evidence pointing to determinism, however—each time we drop a pencil and it falls to the ground with an acceleration of 9.8 meters per second, we are confirming that events in the world operate according to unchanging laws—in this case, the laws of gravity. Each time we put a pot of water on the stove to boil, we are implicitly relying on the deterministic principles of cause and effect: the heat causes the hydrogen and oxygen molecules to move at a faster pace. Quantum theorists will be quick to note the inherent unpredictability of subatomic particles, but few would challenge the claim that the world can still be described as operating within a system of laws. The chance that these “laws” of nature might not accurately predict the state of the world is infinitesimally small and can effectively be thought of as zero. Thus, while the world might not fall on the *exact end* of the determined

³ The term “libertarian” is used here separately from its definition in political philosophy.

side of the spectrum, as far as our current scientific knowledge can tell, it is almost indistinguishably close to it.

Free Will

Unfortunately, it seems that our common sense notion of free will is flawed regardless of one's opinion on determinism: free will is impossible in a world that is determined (so that our choices are made for us by the laws of physics, cause, and effect), *undetermined* (so that our choices are selected randomly without relation to our goals and opinions), or anywhere in between (so that some of our choices are determined by laws and some of them selected randomly). Intuitively, we think of free will as our ability to choose whether or not and how to act. However, in the context of the current discussion, that seems like an unrealistic expectation. What we are *really* referring to (or, at least, should be) when we talk about free will is the extent to which we *feel* that we are in control of our decisions and actions. For example, one would feel freer if she indulged in a chocolate ice cream cone because she were craving chocolate than if she did so because someone held a gun to her head and ordered her to lick or die. As with determinism, it is helpful to think of free will as existing on a continuum (although a separate one), with the feeling of free will on one end and the feeling of complete constraint, or to put it in legal terms, compulsion, on the other (i.e., the gun-to-the-head scenario). We spend most of our time, I would wager, near the "feeling of free will" side, with occasional constraints (i.e., my doctor said I shouldn't eat sweets; I'm too tired to go to the ice cream parlor) pushing us in the opposite direction. In other words, free will is the *perception* of freedom by agents acting in the world.

I say perception, because we may not (and probably do not) have an accurate understanding of the extent to which we are constrained and probably overestimate the amount of autonomy we actually exercise. Although we feel a qualitative difference between eating chocolate ice cream because it tastes good and being forced to do so under the threat of death, that difference may be illusory—especially given what we know about determinism and indeterminism. If we assume that we are living in a deterministic universe, and that *all* of our actions are caused, then we are under an equal amount of compelling force in both cases (the craving vs. the gun). If we assume that we are living in an *indeterministic* universe, the compelling forces are still equal: equally random. Any assumption in the middle of the spectrum would be similarly unfulfilling.

The distinction we sense between the gun and the craving comes from the differential saliency of the compelling forces. This, I argue, is at the root of what makes us feel as though we have free will; we feel that we are free when we are unaware of the compelling forces at work (because they are subtle and, in many cases, unknown), and we feel that we are less free when we are aware of the compelling forces, as when we are staring down the barrel of a shot gun. One could empirically test this hypothesis by having participants read multiple scenarios in which people make decisions based on compelling forces that vary in saliency. Participants would rate how "free" the person is in each scenario. I predict that their freedom ratings would decrease as the saliency of the compelling forces increased.

Compatibilism

Given this new, proposed definition of free will, let us consider compatibilist arguments. As mentioned earlier, compatibilists believe that it is plausible for us to have a meaningful kind of free will in a deterministic universe. Although compatibilist arguments vary, a particularly helpful summary by Gifford (2007) characterizes the most common, modern compatibilist argument as follows:

Compatibilists hold that free will exists as long as the agent is not subject to outside coercion. As long as our actions are a product of our own reasons for taking those actions, as long as they spring from our personal beliefs and preferences, even though those reasons, beliefs, and preferences are a product of a deterministic universe, we have free will. (p. 273)

When compatibilists admit that the “reasons, beliefs, and preferences are a product of a deterministic universe,” they are essentially agreeing with my contention that we are always under the influence of compelling forces. However, they claim that we have free will “as long as the agent is not subject to outside coercion.” These two statements are contradictory. As I have demonstrated, regardless of whether the world is determined, undetermined, or anywhere in between, we are *always* subject to outside coercion; we just *feel* that we are not when the forces that are compelling us are unknown or lack saliency. It seems that the compatibilists fall prey to this ever-convincing illusion. Their argument could be accurately restated as such: one is free as long as the compelling forces that are motivating her decisions are not salient, e.g. we are free when we feel that we are free. As persuasive as our feelings of free will may be, they are illusory, and *feeling* that we are in control of our decisions and actions is not the same as actually *being* in control.

Unfortunately, as Greene and Cohen (2004) pointed out, there is no way around the fact that intuitive free will is inherently libertarian, not compatibilist. And as we have seen, libertarian doctrine does not stand up to the test of science. Thus, with both libertarianism and compatibilism knocked out of the running, we are left with hard determinism and a diminished form of illusory free will.

Section II: Can Neuroscience Inform the Free Will Debate?

Since it has been established that determinism and free will are actually separate (though related) issues, I will consider neuroscience’s relevance to them individually. I have also demonstrated that libertarian and compatibilist arguments leave much to be desired. Even so, many people still cling to libertarian and compatibilist philosophies. Let us examine how neuroscience can add to the free will debate.

Determinism

As discussed earlier, although most of the scientific community agrees that the world is deterministic, the consensus is not unanimous, and there is even more contention about the matter outside of the field. Despite decades of research (and, for those of us who have not witnessed miracles, personal experience) all pointing to the conclusion that every event has a cause and operates according to unchanging, physical laws, some are still left unconvinced. This is not entirely surprising considering how distinctly “ours” the decisions we make seem. While neuroscience cannot *prove* that determinism is true anymore than previous research has, it can lend some pretty convincing supporting evidence.

I hypothesize that neuroscience’s specific brand of evidence will be even more persuasive—especially to the lay public—than previous research has been, because it deals not with the theory of relativity or the continued accuracy of the laws of gravity or thermodynamics, but with the very seat of the human mind. For centuries, it has been easy to put the mind up on a pedestal and claim that it operates by fundamentally different rules from the rest of the world. But if the days of acceptance for Descartes’ proposed “animal spirits⁴” and other similar types of sloppy metaphysics have not yet passed, they are now in their final hoorah. With functional neuroimaging and other emerging neurotechnologies, we are now capable of opening the black box of the brain and peering in, at least to a greater extent than ever before—an ability that will shake the brain’s pedestal, if not knock it down entirely. Although neuroscience cannot revolutionize the free will vs. determinism argument itself, it may revolutionize the way people think about it, as experimental evidence hits closer and closer to home.

For example, by utilizing functional magnetic resonance imaging (fMRI), researchers are able to observe which areas of the brain are active as participants engage in experimental tasks. In one study by Greene, Nystrom, Engell, Darley, and Cohen (2004), participants were scanned while making difficult moral decisions. Greene and his colleagues found that the neural activation varied systematically depending on whether the dilemma was of a personal or impersonal nature. Additionally, depending on the relative activation of the brain centers associated with “cognitive” and “emotional” processing, one could make relatively accurate predictions as to how the participants would respond to the questions being posed. Another experiment by Huettel, Stowe, Gordon, Warner, and Platt (2006) found that differential levels of activation within the lateral prefrontal cortex during a gambling task could predict participants’ preferences for risk taking and general behavioral impulsiveness.

Looking at studies like these, it seems evident that the neural activations researchers are detecting have a causal relationship with the behavior being observed. It also seems clear that it is not an immaterial “soul” that is at work during the decision-making processes, but a very material brain. Furthermore, it is hard to imagine a task that would be more under the “soul’s” jurisdiction than solving a moral dilemma. If the brain is at work solving even this most sacred problem, chances are good (and research points to the conclusion) that the brain

⁴ Descartes proposed that the brain contained “animal spirits” which allowed the mind and the brain to communicate by passing through the pineal gland.

is, in fact, in charge of *all* of our cognitive functions. As these and other studies suggest increasingly mechanistic views of the way the brain works, it is becoming harder and harder to deny that it operates according to the same physical laws as the rest of the universe. As new neuroscientific knowledge pushes conclusions toward the determinism end of the determinism vs. chaos spectrum, we inevitably think about what this means for the free will vs. constraint spectrum. Let us consider how these findings may affect our perceptions of free will.

Free Will

Earlier, I defined free will as a subjective feeling, which depends on the extent to which the actor feels that her choices are constrained by “outside” forces. I hypothesized that it is the saliency of the compelling forces that affects one’s feeling of freedom, with less obvious determinants leading to lesser feelings of constraint. For instance, it is now clear that our genetics and the environment in which we grew up play large roles in the way that we act and the decisions that we make. Nevertheless, because these types of compelling forces are not salient to us on a day-to-day basis, we do not generally feel that we are constrained by them; it is only when pressed to stop and reflect that we discern that our actions are (at least to some extent) products of our biology and upbringing.

Neuroscience will probably play a similar role in our feelings toward free will. Although it is unlikely that we will dramatically change the way we think about our freedom as we shuffle through our daily tasks, when we pause to consider the reality of our condition, we will be forced to admit that we are constrained by the electro- and biochemical mechanisms silently at work within our brains.

As was previously illustrated, fMRI studies have already begun to elucidate the mechanisms by which our brains make decisions, and faster, more accurate technology is sure to arrive in the near future. Other studies have shown even more direct evidence that our feelings of free will are illusory. Libet, Gleason, Wright, and Pearl (1983) published a series of landmark and controversial experiments in which participants indicated when they had come to the conscious decision to execute spontaneous, voluntary movements. This time was compared to the onset of the “readiness potential” associated with the preparation of motor activity, as recorded by electrodes on the scalp. Libet et al. concluded that participants were not conscious of their decisions to make movements until several hundred milliseconds after the first related cortical activity was detected. While the methods of this experiment have been questioned, more recent follow-up studies (Lau, Rogers, & Passingham, 2006; Lau, Rogers, & Passingham, 2007) have shown similar findings, indicating that our brains know that we are going to move before “we” do.

As neuroscience has advanced, our conscious “knowledge” of our intentions has become more questionable, and the causal relationship between the brain and our behavior has become more evident. As the field continues to progress, it is inevitable that the physical, compelling forces involved in our decision-making processes will be more widely understood and, thus, more salient. And just as the presence of a salient compelling force caused greater

feelings of constraint in the gun example, so, too, will the increased salience and knowledge of the biological underpinnings of our minds. It is in this way that neuroscience will affect our notions of free will.

I would like to reiterate, however, that even with this new knowledge, on a regular basis we will continue to feel largely free. We have been groomed by evolution for thousands of years to operate in this way, and so we will continue to do so. Because of this, we will probably still attribute relative freedom to others as well, and make casual judgments of responsibility just as we did before—blaming our friends when they are late for a dinner party and punishing our children for not cleaning their rooms—however, as the salience of determinism limits our conceptions of free will, we will be forced to reassess how we deal with responsibility in more formal contexts. I turn now to examine the most formal of all contexts for responsibility in our society: the penal system.

Section III: Neuroscience and Responsibility

Legal punishment is classically divided into two theories of justification: retributivism and consequentialism. Retributivism is based on the idea of “just deserts”—one should be punished because he *deserves* it, even if punishing him will not lead to any measurable benefits for society. Consequentialism, on the other hand, is based on a more utilitarian view, whereby people should be punished because it keeps them from doing more harm to the public and deters other potentially dangerous individuals from committing future crimes. If, as I hypothesize, the public becomes more sympathetic to a deterministic worldview and no longer believes that people possess what we have previously, intuitively called “free will,” what will this mean for the retributivist and consequentialist justifications for punishment?

Determinism and the Retributivist Theory of Punishment

If determinism is true (e.g., if the universe has operated according to set physical laws since the dawn of time and has proceeded forward in the only way possible given the constraints of those laws) then it seems that no one could ever *deserve* to be punished. After all, we do not blame minors or the mentally ill because we deem that they are not fully in control of their actions and should not be held responsible. Under the assumption of a deterministic universe, we *all* lack control, and if the only justification for punishment is to give the agent “what’s coming to him,” it is inherently unfair to hold *anyone* responsible for his or her actions, as agents who lack control should not be punished.

While it may initially seem counterintuitive to do away with retributive punishment, even under deterministic conditions, the following thought experiment should help to elucidate the argument: If someone were holding a gun to Steve’s head and ordering Steve to shoot Mary, we would not blame Steve because he did not have a choice in the matter. We would all agree that he was compelled to act by outside forces and understand why he felt as though he was at the constrained end of the free will spectrum. If Steve shot Mary because he

had a large tumor in his frontal lobe that was affecting his ability to inhibit his actions (and we were completely confident that it was the tumor that led to the action), after analysis, a reasonable person would say that he was equally as constrained in this case as he was in the previous scenario. Even though it is not quite as obvious to us (as the compelling force is less salient), we understand that he was unable to stop himself from pulling the trigger because of a biological condition out of his control, and, thus, he is equally as blameless as before. Now, let us take this one step further: What if Steve did not have a brain tumor, but just happened to have been born such that the mechanisms in his frontal lobes were slightly and undetectably abnormal? He shot Mary as a result of his innately poor inhibition abilities. Would he still be in the clear? To be consistent, we must answer yes. He is just as constrained by his biology in this case as he was in the last scenario, but the compelling force is even *less* salient than before.

We are all constrained by our biology. Every decision we make is a result of the way our brains have evolved, our individual genetic differences, and the way our environments have molded us. Our brains are physical hunks of matter that are subject to the laws of physics. We did not design our brains, and we certainly did not dictate our genetics, environments or the laws of physics. Therefore, our decisions are out of our control, and if retributivism is our only rationale for punishment and moral responsibility, we should never punish or blame anyone for her actions, and our judicial system needs a dramatic makeover.

However, others disagree with this analysis and claim that there is currently nothing neuroscience can throw at the law that would render it as helpless as the previous argument implies. In a meeting with the President's Bioethics Counsel (2003), legal scholar Stephen Morse laid out his point of view very clearly: All that is required to hold an agent legally responsible is rationality. Neuroscience cannot yet prove that we are not rational actors, and it is very unlikely that it will do so in the future. Therefore, our penal system will remain unaffected. The law does not care if we are "free" in any philosophical, metaphysical sense; as long as the criminal is rational, regardless of *why or how* he is considered to be rational, he is culpable for his crime. Morse pointed out that science will continue to be useful in helping us figure out whether or not people were rational agents at the time that they committed their crimes (i.e. was the defendant sleep walking or temporarily insane), but until neuroscience shows that we are, in general, incapable of rational thought processes, the law is safe.

Morse's argument sounds pretty good on the surface; I agree with his analysis that current and future neuroscientific findings cannot and will not undermine the law in its current framework. However, I join Josh Greene and Jonathan Cohen (2004) in arguing that as society adopts a more deterministic worldview and moves toward the conclusion that humans do not have free will in the intuitive sense that was previously assumed, our moral intuitions will adapt to reflect this knowledge and become incompatible with the penal system in its present form. As Greene and Cohen put it, "The legitimacy of the law depends on its adequately reflecting the moral intuitions and commitments of society." The real trouble for the law is rooted not in the fact that the existing system is unable to handle new neuroscientific developments, but rather, in the more profound problem that the law is no longer consistent with our changing view of free will and therefore, moral responsibility.

Determinism and the Consequentialist Theory of Punishment

In order to remain in line with our moral intuitions, it seems that our justice system must give up on the retributive justification for punishment. However, as long as the consequentialist argument for punishment can hold up to determinism, there may still be a place for the penal system (albeit a modified version) that is consistent with our changing moral understandings. The reader will remember that consequentialist accounts of punishment are not based on backward-looking ideas of just deserts, but on forward-looking ideas of prevention: It is justifiable to punish those who infringe on the rights of other members of society because it will detain current criminals (and thus keep them from doing more harm) and deter future criminals. This explanation of punishment is consistent with determinism because it does not necessitate that the criminal be held morally accountable. He is not punished because he was in control of his decisions and made a poor choice, but because he is a danger to the public. Taking away his rights will ensure that he will not cause any further harm, and his incarceration will serve as a disincentive to others who, without the threat of punishment, may have acted similarly in the future.

But for every point, there is a counterpoint: Goodenough (2004) argued that consequentialist punishment, just like retributive punishment, relies on an assumption of free will and is, therefore, null and void in a deterministic universe. He claims that when we base our justification of punishment on the value of deterrence, we are making an implicit commitment to the idea that the criminal “had the capacity to fully integrate the threat of punishment into [his/her] decision-making calculus, and to act accordingly, i.e. as if he/she had a kind of free will” (pp. 1807).

The problem with Goodenough’s (2004) argument is that one’s ability to incorporate the costs of punishment does not depend on his being free. Our brain includes new factors into its decision-making every day—I feel hungry; considering this new piece of information, my brain weighs the options and decides that the best course of action is to eat in the near future. The threat of punishment can be treated like any other parameter we consider as we go through our days—Steve was jailed for shooting Mary; considering this new piece of information, my brain weighs the options and decides that the best course of action is to not shoot people. The consequentialist argument does not rely on an assumption of free will, but merely an assumption that we are capable of incorporating new information into our decision-making processes, and this just happens to be the brain’s specialty. Therefore, despite Goodenough’s objection, the consequentialist justification for punishment stands up to the exacting test of determinism. With this established, it seems that our penal system should not be entirely tossed out, but refurbished under the assumption that the goal of punishment is entirely consequentialist, rather than retributivist. (See Greene and Cohen (2004) for a more fully articulated version of this argument.)

Summary and Conclusion

The present paper explored the extent to which neuroscientific findings can inform the free will debate. I argued that neuroscience’s most direct contribution to the argument will be

convincing support for determinism that resonates with the public in a way that no scientific evidence ever has before. Additionally, while this may alter our views of free will theoretically and affect the way we think about moral responsibility and our legal system, we will most likely feel as free on a day-to-day basis as ever before. Although it will take time and work for our justice system to reflect our new understanding of responsibility, it is possible to retain a formal system of punishment in a deterministic world.

My interpretation of the free will vs. determinism debate may seem, to some, unfulfilling, incomplete, or just plain wrong. But regardless of one's opinions on the topic, it is important that society keep one eye on new research and one eye on its implications. While the answer to the free-will debate may still elude us, we must continue to discuss this age-old question any time new evidence emerges, as our concepts of moral responsibility rest on the answer's implications. It is important that our justice system continues to reflect society's moral intuitions, wherever they may lead us; as our self-knowledge changes, so must the law.

References

- Churchland, P. S. (2005). "Moral decision-making and the brain" In J. Illes (Ed.), *Neuroethics: Defining the Issues in Theory, Practice and Policy* (pp. 3-16). New York, New York: Oxford University Press.
- Dennett, D. (2003). *Freedom Evolves*. New York, New York: Viking.
- Gifford, A. (2007). The knowledge problem, determinism, and *The Sensory Order*. *The Review of Austrian Economics*, 20, 269-291.
- Goodenough, O. (2004). Responsibility and punishment: whose mind? A response. *Philosophical Transactions of the Royal Society of London B* (Special Issue on Law and the Brain), 359. 1805-1809.
- Greene, J.D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict in moral judgement. *Neuron*, 44, 389-400.
- Huettel, S. A., Stowe, C. J., Gordon, E. M., Warner, B. T., & Platt, M. L. (2006). Neural signatures of economic preferences for risk and ambiguity. *Neuron*, 49, 765-775.
- Hume, D. (1739). *A treatise of human nature*. (L. A. Selby-Bigge, Ed.). Oxford: Clarendon Press. (Original work published 1967)
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*, 106, 623-642.
- Nahmias, E. (2006). Folk fears about freedom and responsibility: Determinism vs. reductionism. *Journal of Cognition and Culture*, 6, 215-237.
- Nichols, S. (2004). The folk psychology of free will: fits and starts. *Mind and Language*, 19, 473-502.

- Nichols, S. (2006). Folk intuitions on free will. *Journal of Cognition and Culture*, 6, 57-86.
- Overbye, D. (2007, January 2). Free will: Now you have it, now you don't. *The New York Times*, pp. F1.
- Rosen, J. (2007, March 11). The brain on the stand. *The New York Times Review*, 11.
- Van Inwagen, P. (1982). The incompatibility of free will and determinism. In G. Watson (Ed.), *Free will* (pp. 46-58). New York, New York: Oxford University Press.
- Wegner, D. M. (2002). *The illusion of the conscious will*. Cambridge, MA: MIT Press/Bradford.
- Youngsteadt, E. (2008, April 14). Case closed for free will? *ScienceNOW Daily News*. Retrieved April 25, 2005, from <http://sciencenow.sciencemag.org/cgi/content/full/2008/414/3?rss=1>