Published in *Cognitive Systems Research* 34-35, Special Issue on Philosophical Approaches to Social Neuroscience (2015): 54-70.

**Will Retributivism Die and Will Neuroscience Kill It?[[1]](#footnote-1)[[2]](#footnote-2)**

**Iskra Filevaa and Jonathan Tresanb**

*aDepartment of Philosophy, University of Colorado, Boulder UCB 232, Boulder, CO 80309 United States*

*bDepartment of Philosophy, University of Rochester, 532 Lattimore Hall, Rochester, NY 14627-0078, United States*

**Abstract:** In a widely read essay, "For the Law, Neuroscience Changes Nothing and Everything," Joshua Greene and Jonathan Cohen argue that the advance of neuroscience will eventually result in the widespread rejection of free will, and with it -- of retributivism. They go on to propose that consequentialist reforms are in order, and they predict such reforms will take place. We agree that retributivism should be rejected, and we too are optimistic that rejected it will be.  But we don’t think that such a development will have much to do with neuroscience – it won’t, because neuroscience is unlikely to show that we have no free will. We have two main aims in this paper. The first is to rebut various aspects of the case against free will. The second is to examine the case for consequentialist reforms. We take Greene and Cohen’s essay as a hobbyhorse, but our criticisms are applicable to neurodeterministic anti-free-willism in general.

We first suggest that Greene and Cohen take proponents of free will to be committed to an untenable homuncular account of agency. But proponents of free will can dispense with such a commitment. In fact, we argue, it is Greene and Cohen who work with an overly simple account of free will. We sketch a more nuanced conception. We then turn to the proposal for consequentialist reforms. We argue that retributivism will fall out of favor not as a consequence of neuroscience-driven rejection of free will, but rather, as a result of a familiar feature of moral progress -- the expanding circle of concern.  In short, retributivism can and must die, but neuroscience will not kill it – humanity will.

**Keywords:** free will, neurodeterminism, legal responsibility, retributivism, consequentialism, Brain Overclaim Syndrome, Josh Greene, Jonathan Cohen, Stephen Morse.

1. **Introduction**

“Man is condemned to be free: condemned, because he did not create himself, yet nonetheless free, because once cast into the world, he is responsible for everything he does.” -- Jean-Paul Sartre, *Existentialism Is a Humanism*

There is a joke about social workers that goes something like this: “Two social workers are walking down the street. They hear moaning and cries for help coming from the nearby alley and go to see what is going on. A man with a face covered in blood, obviously beaten up badly, is lying on the ground. One of the social workers turns to the other one and says, ‘The person who did this really needs help.’” Though this anecdote is, undoubtedly, a caricature of social worker attitudes toward victims and perpetrators, it can nonetheless be said to arise from a not uncommon sentiment – the fear that “experts” on human behavior tend to carry the task of explaining a criminal act by an appeal to causal factors independent of the will so far as to effectively deny the role of free will and portray all perpetrators as victims of their psychological make-up and circumstances.[[3]](#footnote-3)

That this type of fear is not uncommon can be gleaned from the fact that politicians use it in an attempt to win votes. In 1968, George Wallace, running for president as an independent party candidate, declared:

If a criminal knocks you over the head on your way home from work, he will be out of jail before you’re out of the hospital (…) But some psychologist will say, well, he’s not to blame, society is to blame. His father didn’t take him to see the Pittsburgh Pirates when he was a little boy. (Beckett, 1997, p. 34).

In his detailed account of punitive practices in the US, Joseph Margulies writes:

By the end of the decade [the 1980s], the Republican and Democratic positions on crime were nearly indistinguishable. The Democratic platform of 1988 abandoned the now heretical suggestion that crime could be caused by social conditions and pledged an aggressive role for the federal government in controlling lawlessness. (Margulies, 2013, p. 97).

Historically, on the part of the wider public, denials of free will have been met with reactions ranging from fear to ridicule. Perhaps all this is about to change. In “For the Law, Neuroscience Changes Nothing and Everything,” Joshua Greene and Jonathan Cohen reckon that the two questions in our title must be answered in the affirmative. Neuroscience, they contend, will put an end to retributivism in legal adjudication by showing that the notion of just deserts, which forms the cornerstone of the retributivist doctrine, is based on empirically untenable ideas of free will. (Greene and Cohen, 2004, pp. 1775-17785).

The proposal is radical. Greene and Cohen are not just anti-retributivist. Their anti-retributivism follows a deeper rejection of free will. The claim is not that a special group of defendants – adolescents, patients with brain damage, people with behavioral addictions, etc. – may lack the freedom necessary for legal responsibility. Nor is it that all of us may lack it on a particular occasion, when we do something absent-mindedly, say, or in the heat of passion. Rather, the idea is that *no one* *ever* chooses freely what to do.

Yet, the proposal is not new. The view defended by Greene and Cohen is but the latest permutation of a significant though never culturally dominant strand of thought according to which concepts such as freedom and responsibility have no role to play in a scientific worldview and ought to be eliminated from, *inter alia*, legal practices.[[4]](#footnote-4) The twist is that a new kid on the block, neuroscience, is now enlisted in the fight against free will.

This, however, is not a minor addition, according to the two authors. Neuroscience, they claim, will at last enable anti-free willist, anti-retributivist views to achieve supremacy in practice. Popular doubts about free will, they tell us, have thus far been relegated to the back bench in the theater of ideas, because the mind has remained a black box. This has allowed believers in free will to use it as a “donkey on which to pin dualist and libertarian intuitions.” (Greene and Cohen, 2004, p. 1781). Neuroscience will finally turn the black box into a “transparent bottleneck.” (ibid). The metaphor is a reference to the way in which all the different causal influences – genes, physical condition, social factors, and so on – impact behavior: all of these forces must, ultimately, pass through the brain’s “bottleneck” and emerge as features of brain states on the other end. Very soon, Greene and Cohen suppose, we’ll have an up-close view of the activity in the “bottle” in the form of pictures from high-resolution scanners. And not even the staunchest free will defender will be able to stand her ground when that happens. For, they say, it is one thing to resist a general philosophical argument against free will, and quite another to keep supporting the thesis that free will exists in the face of “images of the face of brain structures involved [in human action] and equations that describe their function.” (ibid).

It would be difficult to deny that neuroscience has breathed new life into old ideas. Neurodeterminism, as some have called the view advocated by Greene and Cohen,[[5]](#footnote-5) has gained significant traction in recent years. Greene and Cohen join a big chorus of authors, with expertise ranging from neuroscience to social psychology, who have recently argued that free will as we know it is some sort of illusion: Sapolsky (2004, pp. 1787-1796),[[6]](#footnote-6) Eagleman (2012), Wegner (2002), Gazzaniga (2011), Ramachandran (2012), to mention a few. The debate among scholars has long spilled over into the popular media. Thus, some time ago, *The Economist* began an editorial on neuroscience and free will thusly: “Genetics may yet threaten privacy, kill autonomy, make society homogeneous and gut the concept of human nature. But neuroscience could do all of these things first.”[[7]](#footnote-7) Neither is *The Economist* the only popular publication to echo this sentiment: “How Physics and Neuroscience Dictate Your ‘Free Will” (Koch, 2012), “Free Will Could Be the Result of Background Noise in the Brain, Study Says” (Molloy, 2014), “Free Will is a Cognitive Illusion” (Bering, 2013) are a few representative recent titles. “Your brain made you do it” is fast becoming the new mantra.[[8]](#footnote-8) Perhaps what we are witnessing is an early stage of the tidal shift in public opinion predicted by Greene and Cohen.

What is “neurodeterminism”? We say more about this shortly, but the basic idea is that determinism is true for all practical purposes, even if not strictly true (e.g,. due to indeterministic quantum events). Greene and Cohen attempt to make the view attractive not only in theory but in practice, by pacifying the fear associated with the denial of free will. They have good news for those who, like the George Wallace voter, are anxious about what Dennett once called the problem of “creeping exculpation.” Even if there’s no free will it doesn’t follow that punishment is unjustified. For, they remind us, retribution and deterrence are distinct justifications for free will, either one of which might be sufficient on its own. And it is only the former that requires free will. Since we aren’t free, retributive justice must go. But when it comes to punishment, deterrence may still carry the day, and deterrence does not require freedom. They write:

People may grow completely used to the idea that every decision is a thoroughly mechanical process, the outcome of which is completely determined by the results of prior mechanical processes. What will such people think as they sit in their jury boxes? Suppose a man has killed his wife in a jealous rage. Will jurors of the future wonder whether the defendant acted in that moment *of his own free will*? Will they wonder if it was *really him* who killed his wife rather than his *uncontrollable anger*? Will they ask whether he *could have done otherwise*? Whether he really *deserves* to be punished, or if he is just a victim of unfortunate circumstances? We submit that these questions, which seem so important today, will lose their grip in an age when the mechanical nature of human decision-making is fully appreciated (Greene and Cohen, 2004, p.1781 )

And they offer a prediction:

The law will continue to punish misdeeds as it must, for practical reasons, but the idea of distinguishing those who are truly deeply guilty from those who are merely victims of neuronal circumstances will, we submit, seem pointless. (Greene and Cohen, 2004, p. 1781).

Greene and Cohen, then, offer two parallel lines of thought, one argumentative, and one – predictive. Argumentatively, they say that neuroscience does indeed favor certain traditional philosophical positions on free will, including neurodeterminism and the rejection of free will itself. They suggest that rejecting free will underwrites the rejection of desert, leaving consequentialist considerations dominant in the justification of punishment. A consequentialist turn calls for a transformation of our practices of criminal punishment or at least current practices in the United States, since those practices are largely shaped by non-consequentiaist, retributivist considerations, which rely crucially on the existence of desert and thus free will.

Predictively, Greene and Cohen suggest that neuroscience will in fact contribute to a wider appreciation of neurodeterminist, anti-free willist arguments, and that that will in turn lead to the rejection of desert, and thus retributivism. Consequentialism will thus gain ascendancy, not just in theory, but, eventually, in practice – in the practice of criminal punishment in particular. Since their arguments imply the appropriateness of these changes, their predictive hypothesis is, from their point of view, an optimistic one.

We have two main aims in this paper. The first is to rebut various aspects of the case against free will. We take Greene and Cohen’s widely read essay as a hobbyhorse because it is the clearest articulation of the ideas we want to take on, but our criticisms are applicable to neurodeterministic anti-free-willism in general. We suggest that Greene and Cohen are too quick to assume that neuroscience vindicates neurodeterminism. They take proponents of free will to be committed to an untenable homuncular account of agency. But proponents of free will can dispense with such a commitment. In fact, we argue, it is Greene and Cohen who work with an overly simple account of what free will might be. We sketch a more nuanced conception.

We then turn to Greene and Cohen’s optimistic hypothesis. Here, we express some sympathy. We agree that retributivism’s best days (in terms of social power) are behind it, and that this is a good thing. But we think this is good not because there’s no free will. And we doubt that the waning of retributivism will be largely due to the neuroscientific revelation of a transparent bottleneck in our skulls. Rather, we suggest, if Greene and Cohen’s hoped-for change does occur, it will be due to a familiar feature of humanity’s moral progress not essentially linked to neuroscientific progress: the expanding circle of concern. Retributivism’s powerful influence on contemporary criminal punishment bespeaks an incompletely expanded circle, with those taken to deserve punishment left on the outside. The blithe toleration and even righteous approval of abhorrent prison conditions provides evidence for this. Nevertheless, it seems plausible that the circle will eventually expand here too. If it does, the result will be the waning of retributivist influences on legal practice prophesied by Greene and Cohen. But this will no more be due to neuroscience than the (relative) waning of racism has been due to biological discoveries about the nature of race. Rather, the crucial step will be, roughly, the social mobilization of compassion, in which neuroscience will play a minor role at best. In short, retributivism may die, but if it does, neuroscience won’t kill it: humanity will.

1. **Skepticism about freedom in its place**

Is there evidence for determinism?Suppose there were. What would be the consequences for legal practices? One response would be to admit that human actions are causally determined but deny the implications for legal practices that Greene and Cohen seek to draw. This is, roughly, the line taken by Stephen Morse (Morse 2006a, 2006b, 2013).[[9]](#footnote-9) In a series of essays, Morse has argued that for the law, neuroscience changes nothing. This is because legal doctrine, on Morse’s reckoning, is perfectly compatible with everything that neuroscience may throw its way, and it is likely to remain compatible with future science. To conclude that a person is not responsible for an action from the fact that a causal mechanism has played a role – even a significant role – in that person’s decision-making is, according to Morse, to commit the “fundamental psycholegal error.” For it always remains up to legislators to say whether and to what extent a particular bit of evidence from neuroscience is *relevant* to the finding of guilt. And according to the law as we currently have it, Morse contends, an agent is responsible for an offence if the agent is rational and acts intentionally, with the appropriate *mens rea*. Neuroscience cannot show that these criteria are not met by providing evidence to the effect that rationality and mental states are caused, since what the law cares about is whether a person is rational and acts intentionally, not whether rationality, intentionality, and the actions they produce are or aren’t caused. The law, that is to say, is compatibilist. Morse concludes that for the law, neuroscience changes nothing. Somewhat tongue in cheek, he calls the hope that neuroscience will revolutionize law “brain overclaim syndrome,” a condition which, in his view, the likes of him must “treat” with “cognitive jurotherapy.”

In response, Greene and Cohen argue that while the letter of the law may, indeed, allow for compatibilism, most people’s intuitions of justice rely on an incompatibilist idea of free will. Consequently, if neuroscience confirms neurodeterminism, people will no longer intuit that defendants are truly responsible. But neuroscience will confirm that. So for the spirit of the law – if not for law’s letter – neuroscience will change everything.

We doubt this assessment but not for Morse’s reasons. Greene and Cohen seem right to hold that, when it comes to free will, most people have incompatibilist intuitions. Although the matter is complex, empirical evidence seems to point in this direction, though the empirical case is admittedly inconclusive. Eddy Nahmias and collaborators, for instance, have provided data suggesting that people are compatibilists.[[10]](#footnote-10) In one study (out of several), students are first asked whether they believe that the universe is deterministic, and they say “no.” They are then asked whether people would be responsible for acts such as murder in a deterministic universe, and a majority (though not all) say “yes.” (Nahmias et al., 2005). However, Nichols and Knobe, citing a slew of social psychology studies, have persuasively argued that only questions phrased so as to elicit *affect* get a compatibilist response, e.g., questions that ask whether, in a deterministic universe, anyone would be responsible for murder or robbing a bank. When, on the other hand, people are given a detailed description of a deterministic universe and then asked a more abstract question such as whether in that universe it is “possible for a person to be fully responsible for their actions,” a significant majority (86%) say “no” (Knobe and Nichols, 2007, pp. 663-685). So it may be that the compatibilist responses in the affect-eliciting condition are due not to compatibilist intuitions, but, rather, to people’s unwillingness to accept the consequences of their incompatibilist intuitions should the universe turn out to be deterministic.

Nevertheless, even if incompatibilist intuitions are robust and popular, the question remains how we get from there to the rejection of free will. The central argument given by Greene and Cohen is based on a thought experiment and does not involve any evidence from neuroscience.[[11]](#footnote-11) The two authors tell the story of a fictional Mr. Puppet: Mr. Puppet has been carefully designed by a malicious scientist through tight control of genes and environment so that he would commit a murder during a drug deal gone awry. Every single gene in his body has been deliberately selected for this purpose, and every significant event in his life has been scripted. All his teachers, friends, relatives, and enemies have been stooges of the experimenter. But if so, it would be unjust to hold Mr. Puppet responsible. That’s the first step of the argument.

The second step is to extend the conclusion from Mr. Puppet to all of us. For even if it were ceded that Mr. Puppet meets the criteria for diminished responsibility, this fact in itself would not seem to have any wide ramifications since clearly, we are not all playthings in the hands of scientists of ill will. Greene and Cohen anticipate this retort, but they proceed to ask what difference this makes, exactly. What, really, *is* the difference between us and Mr. Puppet? It may seem, initially, that there is a significant difference, since in Mr. Puppet’s case, there are intelligent agents pulling the strings on the forces operating, while no one pulls the strings on *our* actions. But so what? If our actions are determined by causal forces operating on us, does it really matter whether the strings are pulled by an intelligent agent or by blind causality? Surely, Mr. Puppet would be no more responsible than he is in the original scenario if we imagine that everything is as we’ve assumed it – genes, circumstances, etc. – but there is no fiendish experimenter, rather, blind chance has brought about the constellation of those forces. Greene and Cohen write:

What matters is only that these forces are beyond Mr. Puppet’s control, that they are not really *his*. The fact that somebody could deliberately harness these forces to reliably design criminals is an indication of the strength of these forces, but the fact that these forces are being guided by other minds rather than simply operating on their own seems irrelevant, so far as Mr. Puppet’s freedom and responsibility are concerned (Greene and Cohen, 2004, p. 1780).

We begin with Step #1. It would be question-begging for Greene and Cohen to assume that Mr. Puppet *can* be fully controlled through manipulation of genes and environment, however tight. The opponent will insist that it is possible for things to take a course not planned by the scientist, and for Mr. Puppet not to commit murder.

Now, Greene and Cohen admit this possibility, but they imply – without explicitly asserting it – that all the reasons there may be for things to diverge from the scientist’s plans would be reasons that have to do with insufficient knowledge and errors on the part of the scientist, not with some inextinguishable flicker of freedom the manipulated subject possesses. Thus, they write: “Things generally went as planned, but not always. For example, the angry letters written to his dead father were not supposed to appear until he was fourteen, but by the end of his thirteenth year, he had already written four of them. In retrospect I think this was because of a handful of substitutions I made to his eighth chromosome. At any rate, my plans for him succeeded, as they have for 95% of the people I’ve designed. I assure you that the accused deserves none of the credit.” (ibid). (Of course, if Greene and Cohen are right, neither does the scientist deserve any of the credit, but leave that aside.)

Note first that the suggestion that a change in the eighth chromosome can cause Mr. Puppet to send the letters to his angry father an year earlier is very contrived. We bring this up not because we want to quibble about it – it is (obviously) difficult to give a plausible story as to what the scientist may possibly do to ensure that everything goes according to his plans – but rather, because the difficulty in giving a minimally credible story is worthy of notice. It seems more plausible to say – as Greene and Cohen in fact do – that the scientist’s success rate is 95%, not 100%: the former is more plausible, because Mr. Puppet is assumed to be a person, a thinking being, not a robot. Mr. Puppet may, for instance, despite his name, conclude that murder is wrong and refuse to murder anyone. We can stipulate that this is impossible, but then the link to real people becomes more tenuous.[[12]](#footnote-12) Greene and Cohen give us no reason to think that any potential divergence from the scientist’s plans will inevitably have to do with the imperfect nature of the scientist’s knowledge and not with the ineliminable element of freedom.

But be that as it may, we have to cede that Mr. Puppet is, in any event, not *fully* responsible for his actions. A humane justice system would have to see the scientist’s interference as a mitigating factor, if not as a fully excusing condition. The more important question is what this means for the rest of us. This is where we come to Step #2 of the above argument, the claim that really, we are *all* in Mr. Puppet’s position.

One of the premises in Greene and Cohen’s argument for the conclusion that we are all puppets is correct: it is true that *both* intelligent and non-intelligent forces can diminish responsibility. Surely, a mental illness or a tumor in the brain can diminish or even extinguish responsibility, although neither is an intelligent agent. What matters is whether forces extraneous to the agent have *so* diminished his ability to choose freely as to absolve him of responsibility, not whether those forces are intelligent or not. However, in most ordinary cases, our ability to choose freely, while not unlimited, is by no means reduced to the point at which the idea of just deserts becomes untenable. It may seem, at first glance, that our respective reactions to the case of Mr. Puppet and to an ordinary case differ from each other because there is an intelligent agent behind the forces operating on Mr. Puppet, and there isn’t one behind the forces operating on us in typical cases. However, the real reasons for the difference in intuitions lie elsewhere: in most ordinary cases, the causal forces exerting influence on an agent cannot be plausibly seen as seriously diminishing let alone extinguishing his or her responsibility. There is a reason why brain images introduced in court always involve abnormalities – brain damage, arachnoid cysts, etc. – *they are never images of normal brains working normally*. The same goes for citing other risk factors, such as a genetic propensity toward violence, perhaps combined with environmental factors, e.g., having delinquent friends and living in an impoverished neighborhood. But few offenders have serious brain abnormalities and in relatively few cases is there a combination of other serious risk factors that can be plausibly seen as sufficiently explanatory of why a given crime occurred.

Indeed, there are no known risk factors in the case of some crimes. Consider, for instance, the 1924 high-profile case of the murder of Bobby Frank,[[13]](#footnote-13) a case which would later provide inspiration for a number of books, movies, and off-Broadway shows, including Hitchcock’s highly acclaimed film *Rope*. The perpetrators in the case, Nathan Leopold and Richard Loeb, were two well-educated teenagers who’d grown up in luxury and who were, in addition, extraordinarily gifted. Loeb and Leopold had no history of family violence, no delinquent friends, and they had never lived in a crime-ridden neighborhood. None of the usual risk factors could have possibly played a role. Leopold was a child prodigy who spoke multiple languages, while Loeb remains, to this day, the youngest University of Michigan graduate. The two young men had, in everyone’s estimation, brilliant futures awaiting them. However, they began a series of crimes ranging from petty theft to arson, a series that culminated with the cold-blooded murder of an innocent 14-year old boy.

In giving this example, we do not wish to suggest that people are responsible only in “no-risk factor” cases. Rather, we want to show how untenable a general denial of responsibility would be, for such a denial would imply that even Leopold and Loeb could absolutely not help it. That they were bound to kill Bobby Frank. This claim stretches credulity. It is one thing to argue that a particular factor known as a risk factor may have increased a person’s propensity to commit a crime and that it should, therefore, be seen as mitigating that person’s responsibility; it is quite another to suggest that, *regardless of what we know about the presence or absence of risk factors*, we should simply assume that neither a defendant nor anyone else for that matter ever deserves praise or blame for anything, because it is always our brains that “make us”[[14]](#footnote-14) do it. What reason do we have to accept *that* claim?

Presumably, we would accept it if made to reflect on the explanatory power of neuroscience. Greene and Cohen seem certain that neuroscience can yield such results, but we’re skeptical, in light of a point made by Adina Roskies.[[15]](#footnote-15) First, though, some clarification may help.

There are three relevant issues in this ballpark: (i) what sorts of processes are actually involved in the production of human action; (ii) whether those processes suffice for free will; and (iii) the normative/evaluative implications of (i) and (ii), regarding desert and other aspects of moral responsibility, and ultimately for criminal punishment. We’ll begin with (i) and will return to (ii) and (iii) later.

There are three main positions with respect to (i): agent-causal, contra-causal libertarian, and what might be called “determinism for all practical purposes,” but which we’re calling “neurodeterminism.” Proponents of all agree that ordinary causal relations of the sort found throughout the inanimate world are pervasive in neural processes as well. They also agree that that isn’t quite the whole story. But they differ on what more is involved.

Views of the nature of causation vary widely but one feature they share in deterministic accounts is that if some event E at *t* is fully caused, then at some point prior to *t* the entire state of the universe plus the laws of nature entail E.[[16]](#footnote-16) Proponents of agent-causation deny this, at least where the E in question is an action. They hold that people have causal powers – capacities to make things happen – in the sense that *people* are the cause of those happenings. Thus, if Smith dances, her dancing is not uncaused but caused by her. Since her existence is itself caused (originally and persistently), nothing on this picture need be uncaused. Nevertheless, Smith had the capacity to not dance; she could have sat down instead. Had she sat down, that too would have been caused (by her), not uncaused. Nothing about the world prior to her dancing or sitting need have differed for her to do one rather than the other, not even the laws of nature. Agent-causation thus reconciles full causation – the denial of uncaused events – with the rejection of the deterministic assumption noted.[[17]](#footnote-17)

Both libertarians and neurodeterminists reject a role for agent-causation, but differ over the role of uncaused events in the production of action. For libertarians, there are uncaused events playing the role which, by all accounts, exercises of free will play if there are any. For instance, following Robert Kane, they may say that some momentous, impactful life-choices involve genuine indeterminacy.

Neurodeterminists reject both agent-causal and libertarian processes. This doesn’t mean they are strictly determinists, for they may allow such things as quantum indeterminacy, as Greene and Cohen in fact do. But they deny that indeterminacy helps secure free will. This is not just because they deny that randomness can secure freedom. They also deny that the uncaused events correlate even roughly with the alleged occurrence of free choices, taking uncased events, instead, to be distributed throughout physical processes (i.e., uncaused events occur all the time in different places, not simply when people are making a free choice). Neurodeterminism thus differs from determinism in allowing for indeterminacy. It differs from the latter also in being, strictly speaking, non-committal with respect to what happens outside human bodies and brains, for such external processes are irrelevant to the issues at hand (hence the “neuro”).

Greene and Cohen claim that neuroscience confirms neurodeterminism. We deny this. Unfortunately, they don’t even consider the agent-causalist option, treating neurodeterminism and contra-causal libertarianism as the only games in town. But even setting this aside, as Roskies points out, the actual evidence suggests fundamental neural indeterminacy (Roskies, 2006a, p. 115). That is, the most basic *neural* level, far from *seeming* deterministic, *seems* indeterministic. We emphasize “seems” because, of course, that data is as yet neutral between the three positions. Indeed, it doesn’t even rule out full-fledged determinism (i.e., not just “for all practical purposes”), since the most basic neural level isn’t the most basic physical level. The brain is ultimately composed of materials which aren’t essentially neural. Similar neural constituents can be multiply realized by the non-neural, just as two items can be tables but differently constituted. Even tables of the same type are differently constituted – even tables from the same assembly line. Since different non-neural realizations can yield different neural behaviors, apparent indeterminacy at the neural level doesn’t entail indeterminacy full stop. So neuroscience doesn’t disconfirm determinism, and *a fortiori* doesn’t disconfirm neurodeterminism. But neither does it confirm it.

So we doubt that neuroscience will fill in the gaps in Greene and Cohen’s argument. For readers unconvinced by the Mr. Puppet example, however, Greene and Cohen try a second thought experiment. It is this: you are trying to choose between soup and salad. Advanced software allows you to see the neurons pushing for soup in red and those pushing for salad in blue. You watch the movie in slow motion and find the precise point at which one side wins, say the blue neurons out-fire the red ones, and you say that you will have the salad (Greene and Cohen, 2004, p. 1787).

It is not clear what exactly this thought experiment is supposed to show and how it can be expected to persuade the reader that people have no free will. Of course, it is possible that people get “taken” by what some have called the “the seductive allure of neuroscience,”[[18]](#footnote-18) although we strongly doubt that this will happen in an ordinary case involving no abnormality or anything else unusual.[[19]](#footnote-19) But however this may be, the question, of course, is whether there is a good reason to draw an anti-free-willist conclusion, not what conclusion people will actually draw. And we do not think there is a good reason. Since we are only alive for as long as our brains are active, one would expect that there is *some* neural state corresponding to the mental state we are in when we make a choice. If there are any free choices, presumably those too would have corresponding to them brain states or “neural correlates.” What we need is evidence that there is not simply a *correlation*, but *causation* here.

Greene and Cohen suggest that we would be convinced there is causation if we imagined replaying the movie in slow motion and finding the tipping-point at which “the blue neurons in your prefrontal cortex out-fire the red neurons, seizing control of your pre-motor cortex and causing you to say, “I will have the salad, please” (Greene and Cohen, 2004, p. 1781).

These readers remain unconvinced. Putting aside compatibilist responses, what we want to know is precisely whether neurons *cause* us to say anything. Perhaps there are unrecognized but important indeterminacies. Perhaps your salad neurons out-fire the soup neurons because you *choose* the salad over the soup. To simply assume full causality, and that it runs only in the other direction, is question-begging. We need an argument. Have we been offered one?

There is one argument which comes from Daniel Wegner, whom Greene and Cohen cite: Wegner has argued that conscious will does not cause action, and that both thought and action originate from a common unconscious source (Wegner and Wheatley, 1999, pp. 480-492; see also Wegner, 2002). According to Wegner, the perception of free will arises when certain conditions are met, even in cases in which it is quite clear that a given subject has made no choice.[[20]](#footnote-20) One way Wegner set out to demonstrate this was by designing an experiment in which he essentially tricked people, with some degree of success, into thinking that they have done something that was actually done by another person, e.g., stopping a cursor on a screen (Wegner and Wheatley, 1999, pp. 487-489.)

The first thing to note here is that there are reasons to be wary of Wegner’s interpretation of the data. Wegner’s official purpose is to show that the subject in the experiment can be made to believe she has caused an event she hasn’t caused. But according to Wegner’s own account, the subjects have been asked only whether they *intended* to make a stop, not whether they *caused* a stop.[[21]](#footnote-21) A person may well *intend* to do something and observe someone else doing it before she gets a chance to fulfill her intention. In such a case, she will be not be showing herself to be under any illusion of free willing if she responds “yes” to the question, “Did you *intend* to do X?”[[22]](#footnote-22)

But we won’t dwell on this, because there is a more important point to make. Even if Wegner’s experiment succeeded, no interesting conclusion regarding free will follows from it. From the fact that people can be tricked into believing that they have freely done something they haven’t done (an extraordinarily difficult task, in any event – if you read the description of Wegner’s experiment, you will see that he’s gone to considerable lengths in order to obtain the desired result, and even then, the success rate is not all that impressive[[23]](#footnote-23)), it does not follow that everyone is always tricked. To suggest otherwise would be like saying that, because we sometimes misperceive things, we never perceive anything accurately, indeed, that maybe, there is nothing there for us to perceive, and the whole world is an illusion fed to our brains. An argument along those lines has been made, of course, but few believe there is a reason to take it seriously outside of a purely academic discussion on epistemology.

Now, in fairness, Wegner does not claim that conscious will *is* an illusion. He says only that it *could* be an illusion, i.e., that the evidence is consistent with the skeptical hypothesis. Wegner and Wheatley write, “…we can never be sure that our thoughts cause our actions as there could always be unconscious causes that have produced them both. The impression that a thought has caused an action rests on a causal inference that is always open to question – yet this impression is the basis of the experience of will.” (Wegner and Wheatley, 1999, p. 482). Our response is: yes, this is possible. But has Wegner shown that it is a possibility we need to worry about? It’s also possible that a skeptical hypothesis is true. Indeed, much more down-to-earth scenarios are possible, such as that the CIA has an agent in your closet right now. Putting skepticism about freedom in its place can start with putting skepticism, period, in its place.

Ordinary people understand this well enough. Consider the question why it is not difficult to convince reasonable people that severe cognitive impairments, certain mental illnesses, etc. may be excusing conditions, while the idea that everyone is always excused by default has so little traction. Our hunch is this: in the former case, people see *how* the aforementioned causal factors might have conspired to produce a certain action independently of an agent’s will. In the latter case, they do not see how that’s possible. Neurodeterminists tell us no plausible story as to how normal brain states produced by normally functioning brains can be seen as causing behavior in the determinist’s sense. They propose, rather, to substitute images for an argument.

Neurodeterminists sometimes make another suggestion, however. It is this: perhaps, we can find the precise point on the neuro-action movie at which the blue neurons out-fire the red ones and that point can be shown to *precede* the person’s conscious awareness of having made a decision. If so, we would have to say that the decision was not free.

This is, roughly, the conclusion some have drawn from a widely discussed Libet experiment, in which subjects were asked to flex their wrist when they felt like it, and motor activity in the brain known as the brain’s “readiness potential” was found to begin a few hundred milliseconds before subjects were aware of their decisions. (Libet, 1999, pp. 47-57). This suggestion would be very much in line with Greene and Cohen’s proposal, though interestingly, the two authors do not reference Benjamin Libet, possibly because Libet himself was not a neurodeterminist. He argued, instead, that while we don’t have the ability to initiate acts purposefully, we have the power to veto urges to act and even intentions, a power dubbed by V.S. Ramachandran “free won’t” (as opposed to free will).

There is now a lot of philosophical literature on the Libet experiment and the initial excitement with the results has subsided somewhat, in light of the objections raised.[[24]](#footnote-24) We cannot discuss this literature in any detail and will limit ourselves to making two points.

First, in the Libet experiment, people are specifically instructed to go along with an urge. But what if they were instructed to make a conscious decision first and then wait for 20 seconds, say, before acting? Surely, they could have done it. An inability to comply with this type of simple instruction would be evidence of a medical problem. The neurodeterminist, however, can respond to this by saying that, while we *may* be capable of making a conscious decision and then acting on it, not all – or even most – of our actions are *in fact* preceded by a conscious decision. What is worse, she may go on, even those actions that do result from a conscious decision are ultimately driven by unconscious brain processes, because the making of a decision itself is so driven.

The first point does not help the neurodeterminist much, though it may cast doubt on the practice of legal adjudication: if only the actions preceded by a conscious decision are free, then it may follow that certain actions currently considered culpable, for instance, crimes of passion without premeditation, may have to be fully excused, at least unless there is proof the agent formed a conscious decision to act first. But it will not follow from here that no one is ever responsible.

The second point is more important. In approaching it, we will take a step back first. Note that in both the Greene and Cohen salad over soup example and in the Libet experiment, people have no particular reason – whether normative or strong prudential reason – to choose any of the alternatives available. But when we do not have a particular reason to choose one thing rather than another, we frequently go along with whatever urge we happen to have.

Now, one could argue that if we do not have a reason to prefer one thing over another, we cannot make a free choice or, indeed, any choice. We do not need a neuroscience movie to make that argument. There are sufficiently good *a priori* grounds to claim that choice without reasons will be, to borrow an expression of Harry Frankfurt’s, little more than a “volitional spasm.” (Frankfurt, 1988, pp. 177-190).[[25]](#footnote-25) There may seem to be a tension between Frankfurt’s account and a persuasive argument given by Sartre to the effect that to choose is to create oneself, since “creation” may appear to imply no prior reasons; however, the two views are quite compatible: as with any creative act, choice in action need not be arbitrary, and choices, though a matter of self-creation, can still be good or bad. It is just that, Sartre says, we have no essence that determines our choices, and in this sense, we create ourselves through our choices.

Sartre also suggests that this is why we are responsible for making ourselves the people we are. We won’t discuss this issue here, since we are interested more narrowly in legal responsibility. Whatever the truth of one’s responsibility for self-creation in general, for instance, for choosing to pursue a career as a singer rather than an engineer, the main question for our purposes is whether a person is capable of operating within certain legal constraints, for instance, of choosing a legally permissible career (she may, in addition, be responsible for choosing the particular legally permissible career she chooses and responsible for squandering her life if she chooses an unsuitable career, but this is more than we need to demonstrate).

Can a person choose to operate within the legal constraints? In answering this question, let us consider a different version of the soup and salad scenario. You are wondering whether to have a cup of soup or a salad plate. As it happens, there is only one bowl of soup left. You take this as a sign that the soup is good at this café, and you are thinking of taking the soup. Another customer who was previously in line in front of you but checked out and left comes back and says that he has forgotten to get his soup. It turns out he’s paid for the last bowl of soup already. This naturally ends your process of deliberation – you perceive the soup as no longer an option. Even if you had a slight (or a strong) preference for the soup, you will get the salad.

It is quite conceivable that, if someone were to look into your brain at the moment at which you realize that the soup is now the property of another customer and no longer for sale, and if it is possible for the neurons pushing for the soup and for those pushing for the salad to be colored differently, the salad neurons will be seen to out-fire the soup neurons. Indeed, it is quite possible that, upon hearing that the soup belongs to another, you draw the conclusion that you cannot have it – that the soup option has been foreclosed – before you process the information consciously, and you may withdraw your hand from the soup a few hundred milliseconds before you realize, consciously, what you are doing and why. We submit that, intuitively, you are free in making this decision. The fact that the decision is initiated unconsciously, by automatic affective processes, does not itself seem to make it unfree. Our automatic affective processes are, after all, not disconnected from the conscious processes, unless we have brain damage, for instance to the ventromedial prefrontal cortex (see Eslinger and Damasio, 1985, pp. 1731-1741). Barring such damage, the likely explanation of why dropping the soup option from consideration (or even aborting the intention to have the soup) may be initiated before we have made a conscious decision to do so is explained by the fact that we have internalized a certain constraint on behavior – property rights.

But suppose our chooser heard that the soup belongs to another, yet still wanted the soup, so he stole it. Would he still be responsible? Ordinarily, yes. Information to the effect that an object belongs to another serves as a wake-up call for every properly socialized person. Of course, this may not happen on some particular occasion, for instance, a person may be so deeply affected by grief, say, or so focused on an issue she is working on that she may simply not notice what she is doing, i.e., that she is taking someone else’s property (notoriously absent-minded physicist Norbert Wiener comes to mind here). Issues of this sort are more frequently resolved informally, but if such a case reaches a court of law, the defendant can argue that she did not intend to violate anyone’s rights and that either the *actus reus* or the *mens rea* condition on criminal responsibility is absent, i.e., the act was automatic and involuntary or it was voluntary but without the requisite intention. In order for Greene and Cohen’s argument to succeed, they have to show that even when a person is fully aware of the available alternatives and appears both to herself and to others to be choosing one among those alternatives, she is really not making a free choice. But we need a reason to think she isn’t.

The burden of proof here is on anti-free-willists, since theirs is the counterintuitive view. Typically, we make choices for what we consider good reasons, although sometimes, we decide to do something simply because we want to do it, without the belief that it is truly a good idea. We make our choices in light of the foreseeable consequences of our actions, and we are held responsible for those consequences precisely because we have foreseen them and still made the choice we made. Suppose a person decides to steal someone else’s soup. She does this even though she knows the possible consequences, including legal consequences. Intuitively, she has made the choice freely and thereby committed herself to being held responsible for the consequences. Seeing her brain activity at the time of decision will not give us a reason to think she is not choosing freely, unless we discover something unusual or abnormal.

It is, of course, possible that the choice is not as free as it, at first, appears. For instance, someone may have misled our agent into believing that the soup does not belong to anyone, or she may have misinterpreted the claim that the soup has been bought by another due to poor English language skills, or else she may have severe anterograde amnesia and be unable to retain the information that the soup belongs to another for long enough in order for her choices to reflect this bit of information. But if there is no such unusual condition in place, we can assume she has acted freely.

**3. Freedom as a matter of degree and context: beyond “either, or”**

“Yet, now that I recall all the circumstances, I think I can see a little into the springs and motives which being cunningly presented to me under various disguises, induced me to set about performing the part I did, besides cajoling me into the delusion that it was a choice resulting from my own unbiased judgment and my own free will.”

-- Herman Melville, *Moby Dick*, Chapter 1 (soliloquy of Ismael)

Greene and Cohen suggest that those who reject neurodeterminism are moved by the dualist idea of an immaterial homunculus associated with the brain. Perhaps this is true of some opponents of neurodeterminism, but the rejection of neurodeterminism does not depend on the tenability of that (quite obviously untenable) idea. Opponents of neurodeterminism needn’t be dualists and needn’t believe in the existence of a little agent inside of our heads. If the only way of rejecting neurodeterminism *were* the homunculus theory, the discussion will likely have ended a long time ago. However, homuncularism is not the *only* alternative.

Neither libertarianism nor agent causation requires dualism. This is evident with respect to the former: libertarian metaphysics is as such just like neurodeterminist metaphysics, minus some causes. But the point is also true of agent causalist metaphysics, which requires only persons and the possession of causal powers by persons. These are not light commitments, but neither essentially involves dualism. Persons are posited by all parties equally, including neurodeterminists. (One could, of course, object to free will, desert, and retributivist criminal punishment on the grounds that they mistakenly presuppose persisting persons, but this is not Greene and Cohen’s objection.) And although agent causal powers are indeed peculiar, their peculiarity is isolable. Dualism isn’t necessary for such powers; nor, so far as we can see, does it even help them. A peculiar sort of causation might be *just that*, no more requiring some further, dubious metaphysical backing than does ordinary causation. As agent-causal theorists have emphasized, causation is *in general* mysterious, so it’s not clear how much extra metaphysical burden they bear in the end.

Indeed, far from *implying* a homunculus, non-neurodeterministic free will might be *incompatible* with one. Imagine first that there *is* a homunculus, but it is irrational. It stands outside of natural causality but does not have the ability to act for reasons. Will the existence of such a homunculus amount to the existence of free will? Hardly. Free will is systematically connected to the ability to act for reasons. Although commonly taken as a compatibilist idea, this can also be linked to agent-causal accounts. For free will may involve a capacity to act for reasons.

And even if there is a *rational* homunculus, why would such a homunculus be *you*? We know ourselves as embodied beings. As beings that feel pain when injured, thirst when deprived of water, fatigue after the completion of an arduous task. A homunculus outside the laws of nature would not feel any of these things (if it can feel anything at all), and thus, it wouldn’t be *us*, even if it has the ability to act rationally.

Greene and Cohen, however, do not go so far as to suggest that proponents of free will are explicitly committed to the homunculus theory. They suggest, rather, that the commitment in question is implicit:

[i]n our estimation, even people who do or would readily endorse a thoroughly material account of human action and its causes have dualist intuitions. (…) Consider, for example, the following remarks from Jonathan Pincus, an expert on criminal behavior and the brain:

When a composer conceives a symphony, the only way he or she can present it to the public is through an orchestra….If the performance is poor, the fault could lie with the composer’s conception, or the orchestra, or both… Will is expressed by the brain. Violence can be the result of volition only, but if a brain is damaged, brain failure must be at least partly to blame (Greene and Cohen, 2004, p. 1779)

And they go on to say, “To our untutored intuitions, this is a perfectly sensible analogy, but it is ultimately grounded in a kind of dualism that is scientifically untenable. It is not as though there is *you*, the composer, and then *your brain*, the orchestra. You *are* your brain.” (Ibid.)

Interestingly, Greene and Cohen are guilty of the same double-speak. Recall their claim that the neuron battle movie can show how one group of neurons overtakes our premotor cortex and causes us to say, “I’ll have the salad, please.” If we *are* our brains, it is not clear in what sense our brain’s neurons can cause *us* to do anything. Our brains *are* neurons interacting with each other. Should we take the suggestion to mean that our brains are causing themselves to do things? (If so, that sounds like brain free will, and so like free will *simpliciter* for anyone who thinks we *are* our brains). But we note this point only in passing. Perhaps, the double-speak is unavoidable for reasons that have to do with a deep intertwining between ordinary language and our folk psychological intuitions regarding personhood and free will. Or perhaps, more charitably, Greene and Cohen are deliberately adopting the free willist’s vocabulary in an attempt to meet the opponents on the opponents’ own turf.

The more important point is that Greene and Cohen don’t actually show that those who posit incompatibilist free agency must be dualists. Since, as we’ve argued, the burden of proof is on their side, we could end here. But we wish to turn the tables on Greene and Cohen. They accuse proponents of free will of espousing too crude a conception of what people are like (immaterial homunculi). We think they have a crude conception of what *freedom* is like.[[26]](#footnote-26)

What is freedom, and why suppose anyone has it? A paradigm instance of a free action would be an action done for reasons, where the agent performing the action knows what her reasons are and endorses them. Greene and Cohen maintain that even those actions are not free. So imagine an action of this sort: Sally decides to start saving early for her retirement. She has thought things through and, although she would love to spend more money, she restrains herself and puts a sizable portion of her paycheck in a retirement account. Why would this not be a free action?

One may say, well, Sally is reasons-responsive, and no one chooses to be reasons-responsive. If she weren’t reasons-responsive, she’d waste her money. Yes, but no. Sally is reasons-responsive, but in principle, so is everyone who isn’t mentally disabled or somehow impaired. It is, indeed, true, that none of us endow ourselves with the general ablity to be reasons-responsive – our freedom does not extend *that* far. But it does not follow from the fact that we do not give ourselves freedom that we have no freedom. People who do not have the ability to respond to reasons – mentally disabled people, say, or those with serious brain damage – are exempt from responsibility. [[27]](#footnote-27) The rest of us, although we do not choose to be free, are free and, indeed, as Sartre puts it, “condemned to be free.”

Of course, in any individual case, it is possible for us to discover that we have not acted freely. This is the truth in Wegner’s assertion that it may always turn out our choices are not what they appear to be. But note that legal responsibility does *not* require fully autonomous action. Thus, if you kill someone you hate in self-defense, you may mistakenly believe that your desire to defend yourself was your sole reason for acting. Yet, it could be that, if the attacker were anyone other than your arch-nemesis Chuck, you would have rather risked death than killed. But this will not make you legally blameworthy. The law requires you not to kill anyone except in self-defense. It does not require that killing someone in self-defense be *a fully autonomous action*. Thus, the law here presupposes only a minimal capacity: the ability to abstain from doing what’s not legally permitted. It does not presuppose a capacity to refrain from doing what’s illegal *autonomously*.

But do sane, normal adults have even that minimal freedom? Well, what is the reason to think otherwise? The reason cannot be that an act of killing may be initiated by unconscious processes, since an act initiated unconsciously typically becomes conscious later, and we become free to abstain from it; if, by contrast, the act never becomes conscious, for instance, if a person kills another while sleepwalking, it will not even be counted as an intentional action, and s/he will be acquitted.[[28]](#footnote-28) Nor can the reason be that the individual steps in the complex action of killing may not be within the agent’s control, since we do not hold the agentl responsible for those steps – for instance, for grabbing the knife with the right rather than the left hand – but for putting a particular plan in action.[[29]](#footnote-29) Nor, finally, can the reason be that our little neuroscience movie does not show an agent operating apart from her brain’s neurons. To suggest that even when a person does something consciously, in full knowledge of the consequences of the available alternatives, she is still not acting freely would appear plausible only to those already persuaded byarguments against free will, and so can play no role in such arguments.[[30]](#footnote-30)

We suspect that one reason for what we take to be over-quick dismissal of free will by Greene, Cohen, and others similarly-minded is neglect of the fact that freedom is not an either-or matter. Intuitively, we have the ability to make ourselves freer. This is mainly why neuroscience, as well as other disciplines that study human behavior, such as social psychology and behavioral economics, are exciting rather than depressing – we hope that by discovering more about our brain processes or other factors, such as situational factors that can be said to influence behavior, we can increase our freedom, and begin making better and more autonomous decisions. The process of maturation, of course, is to a large extent a process of expanding our freedom, and does not end when we turn 18 or 21 or any other age.

If freedom is a matter of degree, then how much is enough for legal responsibility? There seems to be no general answer to this question; it’s like asking how much milk is enough for a cake – which cake? For whom? On what occasion?

As with the question how much milk is enough, how much freedom is enough might be contextually variable – and not just for the law but generally. For instance, perhaps to have freely chosen your profession or party affiliation, it must be true that you have chosen these things in a highly autonomous manner. If so, then if you have chosen to be a doctor simply because your parents always said you should do that, or picked a party affiliation only because all of your friends vote for that party, then you haven’t made the choices in question freely. Consider a passage on Ibsen’s play *The Lady From the Sea* by Montrose Jonas Moses. Moses writes: “She [Ellida, the main character] shows him [her husband, Dr. Wangel] how their marriage has been a false bargain, on his part to fill her loneliness, on her part, to be the mother of his children. Force of circumstances, and not free will, welded them together.” (Moses, 2013, p. 460). Given the details provided by the play, this seems a plausible – and most certainly intelligible – analysis. But were Ellida to argue in court that she is not to be considered legally married to her husband because force of circumstances, not free will, brought her and her husband together, her defense would be considered absurd.

Call this view “contextualism about freedom.” Consider an epistemic analogy. Noncontextualist conceptions of knowledge tend to yield skepticism, depriving us of an important resource for talking about differences between our epistemic relations to, say, events seen up close in the light of day, those seen at night in the distance, and those not seen at all. Contextualist accounts of knowledge, on the other hand, allow the concept to play a practically useful role in marking these distinctions, without forcing us to an untenable dogmatism.

Similarly, a contextualist account of freedom could allow us to mark legally important distinctions between what we do soberly and deliberately, at the point of a gun, or while sleepwalking. Neuroscience can demonstrate that some particular person lacked freedom on a particular occasion or that certain sorts of circumstances erode freedom to a greater degree than is currently realized, e.g., in immediate responses to unexpected assaults,[[31]](#footnote-31) but neuroscience can’t show that the minimal threshold is *never* met.[[32]](#footnote-32)

If this is so then neurodeterminists’ disaffection from free will can be likened to skeptical rejections of empirical knowledge on the grounds that such knowledge can never be as secure as knowledge of logic or mathematics. If empirical knowledge must satisfy standards that knowledge that 2+2=4 can meet, it will be hard to see that we have any. But it need not. Neither does freedom.

We have suggested various nuances in an account of freedom which Greene and Cohen don’t recognize: connecting agent-causation with reasons-responsiveness, freedom as a matter of degree, and contextualism about ascriptions of freedom. If these nuances aren’t appreciated, we may be left with a hard choice between a simplistic account of freedom – seemingly requiring an immaterial homunculus – and the rejection of freedom altogether. It is, then, not out of a desire to be perceived as reasonable moderates that we suggest kkka middle way is better; rather, we favor this way because the matter just isn’t simple.

**4. Consequentialism for the right reasons**

There is a final question we wish to discuss: Greene and Cohen’s proposal for consequentialist reforms. We are sympathetic with the proposal but not with the reasons for it.

Nearly every normative theorist recognizes that welfare considerations as such have presumptive moral weight. That an action will benefit someone is a moral reason to do it; that it will harm is a moral reason against doing it; and there is a strong presumption that the greater the benefit or harm, the greater the reason in favor of or against. Retributivists don’t deny this. Rather, they say, the presumption can be defeated or overridden by desert. In certain cases, they argue, criminal offenders behave in a way that makes their welfare less intrinsically morally important or even inverts its value. That is not because their welfare now conflicts with that of others, either due to resource competition or harmful behavior; rather, their behavior *directly* makes their welfare count for less. This is an irreducibly backward-looking moral consideration. The dispute between “consequentialists” and retributivists is over whether the presumption in favor of welfare is overcome in this way by (some) criminal behavior.[[33]](#footnote-33)

We put “consequentialists” in quotes because one needn’t be a full-fledged consequentialist to be an anti-retributivist of this sort. One could be moved in one’s imposition of criminal punishment by considerations of welfare effects (on criminals and others) and still posit a variety of nonconsequentialist moral considerations; e.g., permissible partiality to nearest and dearest, self-ownership rights, Rossian prima facie duties, constraints on treating humanity as a mere means, and so on. There are a wide range of nonconsequentialists for whom welfarism vs. retributivism in criminal punishment is a live option. Although Greene and Cohen may also foresee a triumph of consequentialism proper, in the paper which is our focus here their prediction concerns consequentialist considerations in legal punishment practices.

Greene and Cohen take a radical route to the rejection of retributivism – the denial of free will. This is certainly not *necessary*. One could resist the retributivist’s line of reasoning at a later stage. For instance, one could posit free will but deny that anyone ever deserves anything. After all, the question of moral desert is a normative one, and its answer cannot be simply read off a theory of free will. Moreover, one could accept moral desert but reject the substantive retributivist claim about who deserves what. In particular, one could deny that even the most heinous criminals thereby deserve to be punished as they are now. There are plausible grounds of desert other than freedom, such as humanity or sentience. One could thus hold that every human being, including violent offenders, deserves to be treated humanely merely in virtue of being human.

Indeed, a non-retributivist view of desert of this sort may naturally attract those who *reject* free will. Greene and Cohen hastily presume that abandoning free will inevitably leads to abandoning desert altogether. But unloved, abused children, even very young ones, plausibly deserve better. And so do factory farm animals. If so, it isn’t because of the free choices they’ve made. Exercising one’s free will in certain ways doesn’t seem necessary for one’s deserving anything at all. Thus, anti-retributivist opponents of free will such as Greene and Cohen can say, more modestly, that the implication of free will’s non-existence isn’t that no one deserves anything but that *what* one deserves is never affected by one’s exercises of one’s free will (as there is no such thing). They could then say that, like you and me, even the worst criminals deserve better than to have their welfare discounted in the ways favored by retributivists. Retributivists would thus still stand convicted of succumbing to a myth of desert. But it wouldn’t be the myth that there’s any desert at all.

Finally, one can grant not only desert but the particular claims about desert made by retributivists yet deny that those claims bear in the relevant manner on the all-things-considered moral ought claims which properly govern the application of criminal punishment. There are a variety of ways in which this could be done, and only one of those amounts to denying the relevance of desert wholesale. Alternatively, one could posit an asymmetry in desert’s moral relevance, holding that while an action’s resulting in people getting *less* than they deserve makes it morally worse, an action’s resulting in people getting *more* than they deserve does not. One could hold that the sheer significance of the welfare at stake in matters of criminal punishment simply swamps the impact of desert. Or one could hold a mixture of such views.

But we think the crucial welfarist insight about criminal punishment is that the nature of the processes resulting in action cannot influence the moral value of the agent’s welfare in the way suggested by retributivists, *whatever they turn out to be* and thus whatever else they might imply about such things as her desert. That is, the value of an agent’s welfare cannot be influenced by such things as whether neurodeterminism is true or false of her, and if false, whether that’s because there’s agent-causation or libertarian indeterminism. Consider an analogy: whether someone’s *race* directly influences the value of that person’s welfare shouldn’t be taken to depend on detailed matters about the nature of races, e.g., on whether they’re biologically robust or socially constructed. Of course, the processes resulting in action are relevant to such things as whether an agent will repeat an action again or on what is most likely to deter recurrence. But that is to say that the processes in question have a relevance to forward-looking welfare considerations.

Welfarists who take this position can have an easy time of it with issues (i) and (ii) mentioned earlier – the sorts of processes involved in the production of human action and the question whether those processes suffice for free will. They can grant whatever retributivists wish to say about the processes resulting in action, and about whether those involve free will, and still maintain their rejection of retributivism. Theykk can even grant that punishment is sometimes deserved, rejecting instead the inference from “deserved” to “ought to be administered.” They would thus treat concepts such as “what one deserves” as akin to thick concepts such as “cruel,” “dishonest,” or “courageous.” These seem in a way evaluative, but they do not have an obvious conceptual link to all-things-considered moral ought claims.[[34]](#footnote-34) Full-fledged welfare consequentialists, for instance, can say that actions are cruel, dishonest, or courageous, while maintaining that whether they ought to be performed is entirely determined by their total welfare effects. Similarly, anti-retributivists can grant that subjects may deserve punishment without taking that to make a difference to whether anyone ought to be punished.

In short, consequentialist reforms may be a welcome improvement and may help achieve better outcomes for both society and individuals even if we are free enough for the doctrine of retributivism to get a grip. For instance, even though a person may be fully capable of staying within the bounds of the law and be responsible for violating laws in case s/he does violate any, we could make it easier for him or her not to do anything illegal by, for instance, providing therapy to those who have an impulse control problem or by working to mitigate the various risk factors of crime. There may be good reasons to help even when people *can* make it on their own. That we may be able to make it easier for them to succeed could be a sufficient reason to help. After all, this is how things work with help in general. If, for example, you ask me to proofread your manuscript, and I agree, this is not because I believe that you cannot proofread it yourself. (Surely, if I refuse, and *you* don’t proofread it either, you will be responsible for not proofreading it).

Then there are the interests of society. If it turns out that positive reinforcement rather than punishment makes people with psychopathic traits abstain from crimes, then it would be good to substitute positive reinforcement for punishment even in cases in which a person meets the criteria of legal responsibility.[[35]](#footnote-35) If rehabilitation best reduces recidivism, then it may be justified even for criminals who were sufficiently free and thus deserve punishment.

Rehabilitation is also, in general, a more humane type of treatment. A morally evolved version of the doctrine of just deserts would require us to never punish anyone who isn’t guilty, but only a vengeful version of the doctrine would require punishing everyone who is. We can, after all, treat each other better than we deserve, and there may be excellent consequentialist grounds for doing so.

Finally, we worry that Greene and Cohen have an over-intellectualized view about how social changes such as the shift from retributivist to welfaristic motivations in criminal punishment occur. It seems to us fanciful that neuroscience, no matter how advanced, will be the main driver of this change, should it occur. Nor will neuroscience yield “new appreciation of old arguments” (Greene and Cohen, 2004, p. 1775) against free will on a sufficiently wide scale to yield such changes. Our emotive-motivational structures simply aren’t designed to be so strongly influenced by such things.

We suggest a different frame, whereby the amelioration of excessive criminal punishment will be primarily driven by factors other than widespread shifts in metaphysical or scientific ideas. For reasons we’ll get to shortly, such factors may be non-accidentally *accompanied* by those shifts. But it seems more plausible that if the future witnesses these sorts of improvements, it will be another stage in the general decline of aggression and violence documented so ably by Steven Pinker in *The Better Angels of our Nature: Why Violence has Declined*, and primarily a result of similar psychological, social, and historical trends (Pinker, 2011).

Pinker identifies various “inner demons” and “better angels”: psychological factors inclining us to or against violence, respectively. “Revenge” is evidently the most relevant “demon” driving excessive criminal punishment. One of the modulators of revenge is that its target has become truly harmless. Our felt need for retribution tends to ease only if the perpetrator has lost the opportunity to victimize us again, or has lost the motivation to do so, perhaps due to sincere repentance and reformation. For this reason, technological advances favoring welfarism over retributivism may be much more likely to include those which diminish *fear* of crime than those which diminish the sense of criminal free will. For instance, fear of violence may plummet due to such things as ubiquitous reconnaissance and pacification drones controlled by citizens via direct neural interface.[[36]](#footnote-36) If so, the felt need for retribution may ease.

But we think the most important and relevant modulator of revenge is the ‘better angel’ known as *empathy* (Pinker, 2011, p. 573-591).[[37]](#footnote-37) Empathy is, of course, promoted by awareness. But the relevant object of awareness is not what neuroscience provides, the functioning of brains. Rather, empathy tends to be promoted by factors which elicit our capacities for identification and perspective-taking: e.g., vivid experiential awareness of the facial and bodily manifestations of pain, and relatable life-narratives. Just as the burgeoning of anti-war narratives in the 20th century displaced glory with horror as the main association with war, so anti-punishment narratives might displace righteousness with horror as the main association with criminal punishment.[[38]](#footnote-38) (The film *Dead Man Walking* would be a good example.)

Although neuroscientific information does not as such trigger empathy, where there is neuroscientific awareness of agents, one could reasonably expect empathy-eliciting kinds of awareness as well. In general the advancement of science and science-enabled technological means of communication helps increase this sort of awareness. Consider, for instance, the role of smartphone videos, posted for veryone to see on youtube, in eliciting empathy for the victims of police misconduct. Since neuroscience is itself part of these broader developments, it will not be a mere coincidence if it turns out that an increase in awareness of neuroscientific findings correlates with the rise of ameliorative attitudes, including attitudes toward to criminal punishment. But neuroscience would not thereby be a *driver* of such attitudes.

We do think Greene and Cohen are onto something when they suggest that retributivism feeds on deeply held false beliefs about how criminal minds work. But far more pernicious than any ideas about free will, desert, or homunculism, we think, is what Roy Baumeister (1999) has called “the myth of pure evil.”[[39]](#footnote-39) This myth constitutes a stereotype of criminals (or harmdoers generally), over-stating the harm they do, and the extent to which they are stably and unambivalently moved by an attraction to that harm as such. As Baumeister has shown, this myth is widely held cross-culturally, and shapes the worldview of both victims and observers in various subtle ways, including in motivating harsher criminal penalties.

The widespread debunking of *this* myth, we think, would indeed enable significant progress in the welfaristic amelioration of criminal punishment. But neuroscience isn’t the best way of debunking it. The myth revolves around *psychological* features such as awareness (of harm done) and motives (to harm). Since neurodeterminism and the absence of free will are compatible with any psychological features at all, their acceptance can’t itself undermine the myth. Those in the grip of the myth may simply come to see criminals as akin to Terminators, their behavior manifesting “soulless mechanical processes,” rather than as proper objects of welfaristic concern. Seeing criminals as *machines* won’t undermine stereotypes of them as purely evil; seeing them as *psychologically like us* will; and neuroscience is only indirectly relevant to that. Neuroscience, then, will no more help overcome stereotypes about criminals than it has helped overcome stereotypes about women, African-Americans, or homosexuals.[[40]](#footnote-40) The remedy for psychological stereotypes is exposure to *psychology*, and interacting with people from the stereotyped group. Again, neuroscientific progress may correlate with increases in such exposure, but isn’t yet a form of it.

We suspect that progress in expanding the circle of concern will continue, although of course there’s no guarantee. And we think this progress will increasingly encompass criminals, and ultimately transform criminal punishment in a more welfaristic direction. Such a change, if it occurs, will undoubtedly be accompanied by a shift away from retributivist ideas. But we think this shift will be more of a symptom than a cause. The deeper cause will be that when people get to know people, they also get to care more about them. And when they care more, they are more compassionate and less vindictive. And compassion is a powerful anti-dote to retributivism.

**References**

Baertschi, B. and Mauron, A. (2011). Genetic Determinism, Neuronal Determinism, and Determinism Tout Court. In: *The Oxford Handbook of Neuroethics*, eds. J. Illes and B. L. Sahakian. Oxford: Oxford University Press, 151-160.

Baumeister, R. (1999). *Evil: Inside Human Violence and Cruelty*. New York: Holt Paperbacks.

Beckett, K. (1997). *Making Crime Pay*. Oxford: Oxford University Press.

Bering, J. (2013). Free Will is a Cognitive Illusion. *Big Think*, December 22, 2013, accessed October 3, 2014, http://bigthink.com/in-their-own-words/free-will-is-a-cognitive-illusion

Chisholm, R. (1964). Human Freedom and the Self. The Lindley Lectures, University of Kansas.

Churchland, P. (2006). Moral Decision Making and the Brain. *Neuroethics*, ed. J. Illes. Oxford: Oxford University Press, 3-16.

Clarke, R. (2004). *Libertarian Accounts of Free Will*. Oxford: Oxford University Press.

Eagleman, D. (2012). *Incognito: The Secret Lives of the Brain*. New York: Vintage Books.

Eslinger, P. and Damasio, A. (1985). Sever Disturbance of Higher Cognition After Bilater Lobe Ablation: Patient EVR. *Neurology* 35: 1731-1741.

Frankfurt, H. (1998). Rationality and the Unthinkable. In: *The Importance of What We Care About*. Cambridge: Cambridge University Press,177-190.

Gazanigga, M. (2006). Facts, Fictions, and the Future of Neuroethics. In: *Neuroethics*, ed. J. Illes Oxford: Oxford University Press, 141–148.

Gazanigga, M. (2011). *Who’s in Charge?: Free Will and the Science of the Brain*. New York: Ecco.

Glannon, W. (2011). Brain, Behavior, and Knowledge. *Neuroethics* 4: 191-194.

Greene, J. and Cohen, J. D. (2004). For the Law, Neuroscience Changes Nothing and Everything. *Philosophical Transactions of the Royal Society of London B*, (Special Issue on Law and the Brain), 359: 1775-17785.

Kaposy, C. (2009). Will Neuroscientific Discoveries about Free Will and Selfhood Change Our Ethical Practices. *Neuroethics* 2: 51-59.

Kiehl, K. (2014). *The Psychopath Whisperer*. New York, NY: Crown Publishing.

Knobe, J. and Nichols, S. (2007). Moral Responsibility and Determinism. *Nous* 41: 663-685.

Koch, C. (2012). How Physics and Neuroscience Dictate Your ‘Free Will.’ *Scientific American*, May 7, 2012, accessed October 3, 2014, http://www.scientificamerican.com/article/finding-free-will/.

Libet, B. (1999). Do We Have Free Will? *The Journal of Consciousness Studies* 6: 47-57.

Margulies, J. (2013). *What Changed When Everything Changed: 9/11 and the Making of National Identity*. New Haven: Yale University Press.

Mele, A. R. (2008). Free Will and Neuroscience. In: *Free Will and Luck*. Oxford University Press, 31-48.

Mele, A. R. (2004). Volitional Disorder and Addiction. In: *The Philosophy of Psychiatry*, ed. J. Radden. Oxford and New York: Oxford University Press, 78-88.

Menninger, K. (1968). *The Crime of Punishment*. New York: Penguin Books.

Molloy, A. (2014). Free Will Could Be the Result of Background Noise in the Brain, Study Suggests. *The Independent*, June 21, 2014, accessed October 3, 2014 http://www.independent.co.uk/news/science/free-will-could-be-the-result-of-background-noise-in-the-brain-study-suggests-9553678.html.

Moody-Adams, M. (1991). On the Old Saw That Character Is Destiny. In: *Identity, Character, and Morality: Essays in Moral Psychology*, eds. Owen Flanagan and Amelie Rorty. Cambridge, MA: MIT Press.

Morse, S. (2006a). Brain Overclaim Syndrome and Criminal Responsibility: A Diagnostic Note. *Ohio State Journal of Criminal Law* 3:397-411.

Morse, S. (2006b). Moral and Legal Responsibility and the New Neuroscience. In: *Neuroethics: Defining the Issues in Theory*, ed. J. Illes. Oxford: Oxford University Press, 33-50.

Morse, S. (2013). Brain Overclaim Redux. *Law and Inequality: A Journal of Theory and Practice* 31: 509-534.

Moses, M. J. (2013). *Henrik Ibsen: The Man and His Plays*, London, UK: Forgotten Books.

Murray, D. and Nahmias, E. (2014). Explaining Away Incompatibilist Intuitions. *Philosophy and Phenomenological Research* 88: 434-467.

Nahmias, E. et al. (2005). Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility. *Philosophical Psychology* 18: 561-584.

Nahmias, E. et al. (2006). Is Incompatibilism Intuitive? *Philosophy and Phenomenological Research* 73: 28-53.

Nahmias, E., Coates, J., and Kvaran, T. (2007). Free Will, Moral Responsibility, and Mechanism: Experiments on Folk Intuitions. *Midwest Studies in Philosophy* 31: 214–242.

Nahmias E. and Murray, D. (2010). Experimental Philosophy on Free Will: An Error Theory for Incompatibilist Intuitions. In: *New Waves in Philosophy of Action*, eds. J. Aguilar, A. Buckareff, and K. Frankish. Basingstoke: Palgrave-Macmillan, 189-215.

O’Connor, T. (2000). *Persons and Causes: The Metaphysics of Free Will*. Oxford: Oxford University Press.

Pardo, M. and Patterson, D. (2011). Minds, Brains, and Norms. *Neuroethics* 4: 179-190.

Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford University Press.

Pinker, S. (2011). *The Better Angels of our Nature: Why Violence has Declined*. New York: Viking Penguin.

Ramachandran, V. (2012). *The Tell-Tale Brain: A Neuroscientist’s Quest for What Makes Us Human*. New York: W.W. Norton and Company.

Roskies, A. (2006a). Can Neuroscience Resolve the Issue about Free Will? In: *Moral Psychology Vol. 4: Free Will and Moral Responsibility*, ed. Walter Sinnott-Armstrong. Cambridge, MA: MIT Press, 103-126.

Roskies, A. (2006b). Neurosceintific Challenges to Free Will and Responsibility. *Trends in Cognitive Sciences* 10: 419-423.

Sapolsky, R. (2004). The Frontal Cortex and the Criminal Justice System. *Philosophical Transactions of the Royal Society of London* 359: 1787–1796.

Scurich, N. and Shniderman, A. (2014). The Selective Allure of Neuroscientific Explanations. *PLOS*, September 10, 2014, doi: 10.1371/journal.pone.0107529

Sifferd, K. (2011). Neuroethics. In: *Encyclopaedia of Human Behavior* 2nd Edition, ed. V. Ramachandran. Amsterdam: Elsevier.

Sinnott-Armstrong, W. and Nadel, L. (Eds.) (2010). *Conscious Will and Responsibility: A Tribute to Benjamin Libet*. Oxford: Oxford University Press.

Skinner, B. F. (1971). *Beyond Freedom and Dignity*. New York: Bantam Vintage.

Strawson, P. (2008). Freedom and Resentmnt. In: *Freedom and Resentment and Other Essays* New York: Routledge.

Trevena, J and Miller, J. (2010). Brain Preparation Before Voluntary Action: Evidence Against Unconscious Movement Initiation. *Consciousness and Cognition* 19: 447-456.

Väyrynen, P. (2013). *The Lewd, the Rude, and the Nasty*. Oxford and New York: Oxford University Press.

Vincent, N. (2010). On the Relevance of Neuroscience to Criminal Responsibility. *Criminal Law and Philosophy*, 4: 77-98.

Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Wegner D. and Wheatley, T. (1999). Apparent Mental Causation. *American Psychologist* 54: 480-492.

Weisberg, D. et al., (2008). The Seductive Allure of Neuroscience Explanations. *Journal of Cognitive Neuroscience* 20: 470-477.

Wolf, S. (1987). Sanity and the Metaphysics of Responsibility. In: *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, ed. Ferdinand Shoeman. Cambridge University Press

1. We would like to thank Leslie Marsh and Philip Robbins for their patience and professionalism, and two anonymous referees for comments on an earlier draft. [↑](#footnote-ref-1)
2. The bulk of the work on this paper is by the corresponding author.

*Email address*: iskra.fileva@colorado.edu [I. Fileva] [↑](#footnote-ref-2)
3. Incidentally, Michele Moody-Adams raises an objection along these lines against Susan Wolf’s view on which criminals tend to be victims of either bad upbringing or bad societal influences and may not have received “resources and reasons on which to base self-correction.” See Wolf, 1987, pp. 46-62, 58; also, Moody-Adams 1991, pp. 111-131, 121. Moody-Adams responds by saying: “Indeed, if we call a slave owner or Nazi a victim, then what are we to call the slave or the concentration-camp inmate?” (p. 121). [↑](#footnote-ref-3)
4. See, for instance, Skinner, 1971and Menninger, 1968. [↑](#footnote-ref-4)
5. See, e.g, Baertschi and Mauron, 2011, pp. 151-160. [↑](#footnote-ref-5)
6. Sapolsky expresses the hope that we will one day live in “a world of criminal justice in which there is no blame [but] only prior causes” (p. 1794). [↑](#footnote-ref-6)
7. “The Ethics of Brain Science,” *The Economist*, May 23, 2002, accessed October 5, 2014, http://www.economist.com/node/1143317. [↑](#footnote-ref-7)
8. We found at least three public panels available on the internet with the title “My Brain Made Me Do It,” and a number of others with closely related titles. [↑](#footnote-ref-8)
9. An account similar to Morse’s can be found in Sifferd, 2011. Note also that Gazzaniga too is skeptical about the implictions of neuroscience for legal practice: while he contends that all actions are part of a deterministic system, he argues that responsibility is a social construction and so not the kind of thing one can find in a brain scan. See Gazzaniga, 2006, pp. 141–148. One is also reminded here of Strawson’s now classic paper in which Strawson (2008, pp. 1-28) argued that even if determinism is proven to be true, we won’t be able to call the entire body of our moral practices into question, because we are too deeply committed to those practices. [↑](#footnote-ref-9)
10. See Nahmias et al. 2005, pp. 561-584; Nahmias et al. 2006, pp. 28-53; Nahmias, Coates, and Kvaran, 2007, pp. 214–242; Nahmias and Murray, 2010, pp. 189-215; and Murray and Nahmias, 2014, pp. 434-467. [↑](#footnote-ref-10)
11. The general line of reasoning is familiar. For a more recent version, see Pereboom’s (2014) four-case argument. [↑](#footnote-ref-11)
12. We are reminded of Zak Ebrahim, the son of a terrorist who refused to follow in his father’s footsteps. Ebrahim says: “As I came of age, I began to bend back the bars of bigotry that had imprisoned me for years. It didn’t happen all at once, but little by little, my worldview expanded as I chipped away at every lie my father, and later my abusive stepfather, had instilled in me.” See http://www.huffingtonpost.com/zak-ebrahim/zak-ebrahim-ted-book-terrorists-son\_b\_5811816.html [↑](#footnote-ref-12)
13. See *Illinois v. Nathan Leopold and Richard Loeb*, accessed October 6, 2014, <http://law2.umkc.edu/faculty/projects/ftrials/leoploeb/leopold.htm>. Leopold and Loeb claimed that they wanted to commit the “perfect” crime, i.e., a perfectly motiveless crime. [↑](#footnote-ref-13)
14. It is not entirely clear who the “us” in this statement refers to. We come back to this point later. [↑](#footnote-ref-14)
15. See Roskies, 2006a, pp. 103-126; 2006b, pp. 419-423. For more skepticism regarding the ability of neuroscience to change our moral practices, see Kaposy, 2009, pp. 51-59; and Pardo and Patterson, 2011, pp.179-190; Glannon (2011, pp. 191-194) criticizes Pardo and Patterson on the ground that they do not give the brain its proper due in the production of action. [↑](#footnote-ref-15)
16. We do mean “entail”: if the laws of nature are included, then the relation is not a causal but a logical one. The determinist as we understand her says that every event (with the possible exception of the first) is fully caused, in a way which entails this relation. Agent-causal theorists are not in this sense determinists, though they may say that every event is fully caused. [↑](#footnote-ref-16)
17. For further discussion, see, e.g., Chisholm (1964); Clarke (2004); and O’Connor (2000). [↑](#footnote-ref-17)
18. See Weisberg et al., 2008, pp. 470-477. Others have argued that the “allure of neuroscience” is really “selective,” i.e., that people believe the neuroscientific evidence, when the evidence confirms their prior beliefs. See Scurich and Shniderman, 2014. However, brain images do appear to sway jurors, and prosecutors may attempt to prevent the defense from showing brains scans to the jury. For instance, in the trial of Herbert Weinstein, a 65-year old man who strangled his wife and threw her out the window, in an attempt to make the murder look like a suicide, the defense wanted to present evidence from brain scans showing that Mr. Weinstein had an arachnoid cyst. The prosecution agreed to a plea deal in order to prevent the defense from showing the brain scans, for fear that “simply exhibiting images of Weinstein’s brain in court could sway the jury.” http://www.nytimes.com/2007/03/11/magazine/11Neurolaw.t.html?pagewanted=all [↑](#footnote-ref-18)
19. For what it’s worth, one of us discussed this example with 40 or so students in a *Philosophy and Psychology* class, and not a single one of them seemed to think that the imaginary neuroscience movie described by Greene and Cohen would actually convince them, personally, that the choice of salad over soup is determined. [↑](#footnote-ref-19)
20. The conditions are priority, exclusivity, and consistency: we have a thought of doing something prior to doing it, the thought is consistent with what we do, and we perceive it to be an exclusive cause of what we do. [↑](#footnote-ref-20)
21. Wegner and Wheatley (1999, p. 488) write: “The experimenter explained that the study would investigate people’s feelings of intention for acts and how these feelings come and go. It was explained that the pair were to stop moving the mouse every 30 seconds or so and they would rate each stop they had made for personal intentionality. That is, they would rate how much they had intended to make each stop, independent of their partner’s intentions. The participant and the confederate made these ratings on scales (…) Each scale consisted of a 14-centimeter line with endpoints *I allowed the stop to happen* and *I intended to make the stop*, and marks on the line were converted to percentage intended.” [↑](#footnote-ref-21)
22. We wish to thank Michael Huemer for this point. [↑](#footnote-ref-22)
23. On average, subjects perceive the “forced stops,” i.e., the stops made by the confederate of the experimenter as only *partly* intended. They pick a point approximately in the middle of the scale whose first end point is: “I allowed the stop to happen” and the second end point is: “I intended to make the stop.” [↑](#footnote-ref-23)
24. For instance, Mele (2008, pp. 31-48) has argued persuasively that the “readiness potential” is not necessarily equivalent to an *intention* to act. Mele suggests that the “readiness potential” may, instead, be just an urge, and that the appearance of an urge to act before one has made a conscious decision to act is no threat to free will. A more recent experiment provides partial confirmation of Mele’s interpretation of Libet’s results: in a study conducted by Trevena and Miller (2010, pp. 447-456), subjects were asked to decide, after hearing a tone, whether or not to tap on a keyboard. Readiness potential was detected regardless of whether or not subjects tapped on the keyboard. For other perspectives on Libet’s results, see Sinnott-Armstrong and Nadel (2010). [↑](#footnote-ref-24)
25. Arguably, the idea goes back to Hume. Pat Churchland (2006, pp. 3-16) has re-iterated the Humean point. She writes: “If I suddenly choose to throw my phone out the window, but had no desire, intention, or inclination to do so, if I had no belief that it was my duty to do so, no belief that it was in my interest to do so, then, as Hume saw, this would be a rather absurd action. This action would certainly *not* be considered the paradigm of a free action.” [↑](#footnote-ref-25)
26. K Vincent (2010, pp. 77-98) has defended a similar point about criminal responsibility. [↑](#footnote-ref-26)
27. There are middle-ground cases, of course. For instance, a person with a severe case of acrophobia may find himself able to get on the roof of a house if his toddler has climbed up there and is in danger of falling down. But this does not mean the acrophobe is not volitionally impaired, and in ordinary circumstances, it may be unreasonable and even unethical to demand of him actions that would require overcoming fear of heights. On this point, see also Mele (2004, pp. 78-88). [↑](#footnote-ref-27)
28. This is what happened in the case of Kenneth Parks, who attacked his in-laws and, upon waking up, turned himself in. See *R v Parks*, accessed October 7, 2014, http://scc-csc.lexum.com/scc-csc/scc-csc/en/item/907/index.do. [↑](#footnote-ref-28)
29. Note that ordinarily, the unconscious processes carry out tasks we want them to. They do this only after we’ve “trained our brains” to react in particular ways. Unconscious processes thus help achieve our goals, including the goal of staying out of trouble. [↑](#footnote-ref-29)
30. There is the separate question whether Greene and Cohen have given an argument that would persuade a determinist who is not a neurodeterminist to become a neurodeterminist. Why stop at neural states rather than go all the way to molecular processes, say? We do not think Greene and Cohen give a good argument for this claim either, but that does not matter for our present purposes. [↑](#footnote-ref-30)
31. See, for instance, Kenneth Simons, “Self-Defense: Reasonable Beliefs or Reasonable Self-Control,” *New Criminal Law* *Review*, 11 (2008): 51–90. [↑](#footnote-ref-31)
32. We don’t here undertake to defend contextualism. This would require a careful investigation of linguistic behavior, and an evaluation of explanatory hypotheses which cite pragmatics rather than semantics in explaining contextual variations in standards guiding the application of “free will”, “free”, “freedom”, and so on. Our aim is merely to suggest possible nuances Greene and Cohen overlook, nuances which may lead them to an overly hasty rejection of free will. Also, as noted in section two, we assume that the law is shaped by ideas about freedom, even where it does not explicitly reference them. [↑](#footnote-ref-32)
33. It is thus not at root a dispute over issues of type (i), regarding the processes resulting in action. Retributivists can be agent-causal theorists, libertarians, or (compatibilist) neurodeterminists. [↑](#footnote-ref-33)
34. See Väyrynen (2013) for a defense of an even more serious disconnect between thick concepts and thin moral concepts such as “ought,” “right,” and “wrong.’ [↑](#footnote-ref-34)
35. There is, in fact, evidence that this is the case. See, for instance, Kiehl, 2014. [↑](#footnote-ref-35)
36. Citizens in a society with such technology would in effect be able to call 911 simply by willing it, and guide an immediate and effective response. On one version, drones could quickly pacify everything within a certain radius of a subject, e.g., by monitoring all activity within that area, detecting and incapacitating violent behaviors, and simultaneously activating and informing a more traditional police response. Abuses would be deterred in the same way 911 abuses are these days. [↑](#footnote-ref-36)
37. Interestingly, Pinker himself is equivocal about vengeance-driven criminal punishment, holding that it has various benefits. For instance, it can help internalize inhibitions against violence, facilitate the transfer of punishment functions to central authorities, and minimize the chances of gaming to which consequentialism-driven systems are susceptible. Note that these are all forward-looking considerations, and so don’t amount to an epistemic justification of retributivism. [↑](#footnote-ref-37)
38. Recently, economist Nassim Taleb has voiced criticisms of Pinker’s book *The Better Angels*. The following link contains three pieces by Taleb, one non-technical discussion, and two articles which require knowledge of statistics: <http://www.fooledbyrandomness.com/longpeace.pdf>, accessed July 17, 2015. Briefly, Taleb claims that catastrophic events such as large-scale wars happen rarely, and that during the "lull" moments (which tend to go on for 100 years), we stop worrying about catastrophes, much like the turkey who’s not afraid of its owner, because he comes and feeds it for many months and never injures it, but who gets killed on Thanksgiving. In other words, according to Taleb, the 75-year long period of relative peace and stability we are currently witnessing is a “statistical illusion.”

We are not particularly bothered by this line of criticism. Taleb’s sort of explanation may be good in many areas, e.g., various issues in of Taleb’s field of specialization – economics, such as our tendency to believe that because house prices have been rising for a long time, they will continue to rise, but the explanation is not good here. The difference is that in certain cases such as house market trends, we may be unable to find systematic causes of major shifts, and it may be that such shifts, even when preceded by long-running trends, are, indeed, be due to randomness. But “statistical illusion” seems to be a rather bad explanation of the fact that the Great Powers are no longer going to war with each other, that millions of parents from the current generation have come to regard corporal punishment as unacceptable, that slavery is all but abolished, democracy is spreading, etc. It could be that all of these trends just *happen* to pull in the same direction, but this is simply not the best explanation. Of course, there may still be a large-scale catastrophic war in the future, but if there is, this will be due to the fact that the weapons we have today are vastly more powerful than they have ever been (and a real catastrophe could be caused even by sheer accident), not because we are as violent as people have always been. In fact, ironically, if there is a big catastrophe in the future, it will wipe out the most peaceful human population that has ever lived on this. The interested reader can find Pinker’s response to Taleb here: <http://stevenpinker.com/files/comments_on_taleb_by_s_pinker.pdf>, accessed July 17, 2015. [↑](#footnote-ref-38)
39. Pinker discusses this under the rubric of the “Moralization Gap”. [↑](#footnote-ref-39)
40. Of course, neuroscience has undermined some such stereotypes, at least among certain intellectuals. We don’t mean to deny any relevance of neuroscience at all. [↑](#footnote-ref-40)