Neuroethics (2018) 11:237–244
https://doi.org/10.1007/s12152-018-9376-5

ORIGINAL PAPER

Free Will, Self-Governance and Neuroscience: An Overview

Alisa Carse · Hilary Bok · Debra JH Mathews

Received: 12 June 2018 / Accepted: 14 June 2018 / Published online: 20 August 2018

Abstract Given dramatic increases in recent decades in the pace of scientific discovery and understanding of the functional organization of the brain, it is increasingly clear that engagement with the neuroscientific literature and research is central to making progress on philosophical questions regarding the nature and scope of human freedom and responsibility. While patterns of brain activity cannot provide the whole story, developing a deeper and more precise understanding of how brain activity is related to human choice and conduct is crucial to the development of realistic, just, and intellectually rigorous models of human agency and moral responsibility. In this special issue, we acknowledge that "free will" and "moral responsibility" are not concepts with which neuroscience can directly engage, and instead focus on self-governance, and the capacities that contribute to self-governance, which are more tractable for scientific investigation and are prerequisites for the presence of moral responsibility.

Keywords Free will · Self-governance · Moral responsibility · Neuroscience · Decision-making

A. Carse
Department of Philosophy, Georgetown University, 222 New North, Washington DC 20057, USA
e-mail: carsea@georgetown.edu

H. Bok
Department of Philosophy, Johns Hopkins University, Gilman 278, 3400 North Charles Street, Baltimore, MD 21218, USA
e-mail: hbok@jhu.edu

D. J. Mathews (✉)
Berman Institute of Bioethics and Department of Pediatrics, Johns Hopkins University, Deering Hall 211, 1809 Ashland Avenue, Baltimore, MD 21205, USA
e-mail: dmathews@jhu.edu

Introduction

Philosophers have argued about the nature of moral responsibility for centuries. While they disagree on many aspects of this topic, most agree on one point: that moral responsibility requires the capacity to reflect on our options, reach a decision, and act on it – that is, the capacity for self-governance. Recently, neuroscientists have been exploring the nature and biological underpinnings of this capacity, and the ways in which it can break down. The philosophical literature on moral responsibility has begun to engage more actively and systematically with this important empirical work.

The work in this special issue reflects the rich promise of collaboration across disciplinary lines. We expect it to be of particular value in illuminating the constituent capacities of self-governance, and in deepening our understanding of how self-governance is developed, challenged, diminished, and enhanced. This, in turn, can help us understand what is involved when an individual seems to be incapable of governing her own actions, what specific capacities she might lack, and on

what grounds we should conclude that she is or is not responsible for what she does.

In this paper, we provide an overview to the area of inquiry, discuss the intersections between the work of philosophers and neuroscientists in this space, and summarize and contextualize the articles in this special issue.

Free Will and Moral Responsibility: Persistent Philosophical Puzzles

Most people believe that when we judge a person's actions to be unfair or callous or praise her for her kindness or honesty, we are expressing a different sort of judgment than when we appraise the size of her feet, or admire her perfect pitch. To judge someone unfair, callous, kind, or honest is not just to attribute a defect or excellence to her, it is to appraise her morally. So too when we judge someone to be morally blameworthy or praiseworthy, or respond to her with condemnation or indignation, or respect, or when we ourselves feel guilt or suffer remorse. But what is it about us as persons that makes us the kinds of creatures who are candidates for moral appraisal? And what grounds our responsibility for the morally significant qualities of conduct or character attributed to us?

According to many philosophers, the answer is broadly this: As persons, we are in crucial ways the authors of our own lives. This is a distinctive and morally significant fact about us. Unlike non-human agents – for instance, dogs or giraffes – we are capable of choosing and acting on the basis of reflective deliberation and the reasons it highlights, of pursuing ends grounded in what we ultimately value, rather than what we merely proximally desire, and of undertaking courses of action in spite of countervailing inclinations and dispositions. We are, that is, not simply bound by the "dictates" of instinct or the impress of impulses, desires, and motivations we happen to have. This is not to deny that we can deliberate poorly, suffer weakness of will, or pursue morally problematic ends. Nor is it to deny our susceptibility to fortune and luck. It is to claim that we are, in key ways, able to make moral choices about the lives we live and the kinds of people we are, and that this fact grounds our nature as morally responsible agents.[1]

Within philosophy, the nature and existence of human freedom and responsibility has long been a source of extensive disagreement [8–10, 11–17, 18–21]. Yet, traditionally, most parties to the debate hold that we are morally responsible only insofar as our wills are free. What it means to meet this condition – and, in particular, whether we meet this condition even if our choices and our actions can be fully explained as the effects of antecedent causes – comprises a longstanding point of contention among parties to debates about the nature of freedom and responsibility. On one side of the debate are "compatibilists," who hold that free will is compatible with causal determinism, on the other "incompatibilists" who hold that it is not.

For many incompatibilists concerned with freedom and responsibility, neuroscientific advances invite a particular kind of anxiety. Even if, as some neuroscientists argue, neuroscience cannot establish the truth of determinism or indeterminism, it can nonetheless provide sturdy inductive evidence that our brains are physical systems, or mechanisms, subject to natural law [22]. If our best neuroscientific theories show the brain to be a mechanistic physical system, the concern for the incompatibilist will be that our choices and actions are the effects of antecedent physical causes. Those incompatibilists who do not find comfort in quantum indeterminacy[2] will find in such theories reasons to abandon the belief in our freedom and responsibility.

[1] There is, of course, a great deal of variation among philosophers concerning when and in what ways people are morally responsible. Our claim here is that many philosophers agree that you need a *capacity* for self-governance in order to be a candidate for morally responsibility, in general. This says nothing yet about how exactly self-governance is understood, or about the relationship between some particular *exercise* of this capacity and a judgment that one is responsible for a particular action, attitude, or character trait. On these questions there is much disagreement. Even philosophers who believe moral responsibility extends to actions or traits that are not, strictly, voluntary, but can be attributed to us – perhaps because, for example, they reflect our judgments about the reasons we take ourselves to have – do not think that this would be true even if we wholly lacked a capacity for self-governance. Rather, they question whether a particular action, attitude, or trait has to be track-able to a particular exercise of this capacity if we are to be held responsible for the action, attitude, or trait. See, for example, Scanlon [1] and Smith [2–4]. On the controversy, see Levy [5–7].

[2] Those incompatibilists who are naturalists but believe in an indeterminate universe face the challenge of showing how randomness provides the will causal leverage – how, in an indeterministic universe, what we do is determined by the will rather than by chance. So there is no easy source of "comfort" here.

Concern about the potential of neuroscientific research to upset our views of freedom and moral responsibility has produced a flurry of dramatic proclamations and dire predictions. Joshua Greene and Jonathan Cohen write, for example, "As more and more scientific facts come in, providing increasingly vivid illustrations of what the human mind is really like…[t]he law will continue to punish misdeeds, as it must for practical reasons, but the idea of distinguishing the truly, deeply guilty from those who are merely victims of neuronal circumstances will…seem pointless."[3]

By contrast, compatibilists reject the view that freedom and moral responsibility are threatened by causal determinism or mechanism. They believe that neither our freedom nor our responsibility require that mental activity be independent of physical causes or natural laws. Neuroscientific findings need not, on this view, pose a threat to freedom or responsibility.

However, neuroscience can help to illuminate a different aspect of freedom of the will. Whatever their views on the compatibility of freedom and determinism, most philosophers who write about freedom of the will agree that an agent is free and responsible only if she can step back and ask herself what she should do, make a reasoned choice among her various alternatives, and act on her decision [9, 10, 24]. We will call this the *capacity for self-governance,* and it is the focus of this special issue. Just what constitutive capacities self-governance entails is an evolving and contested question.

We believe that if we did not have the capacity for self-governance we would not have free will and thus, would not be responsible for our actions. If we could not step back and reflect on our motives, then we would be at the mercy of whichever had temporarily gained the upper hand [25]. If we could not reflect on our motives rationally, then we would not be able to figure out on which one we thought we had most reason to act. And if we could not choose among them and act on our decisions, then our assessment of our reasons' relative strengths and weaknesses would be pointless, since it would not issue in action. If any of these things were true, then our conduct would not reflect our own values, judgments or decisions, and according to many philosophers, we would not be responsible for that conduct.

This is not to deny that we sometimes allow our various desires and motivations to dictate our actions, either by failing to exercise our capacity to evaluate them or by deciding that we should not exercise it. But if we are capable of self-governance, then we cannot explain all such failures as the result of our *inability* to decide for ourselves what we have most reason to do, and to act accordingly. Moreover, if we are capable of stepping back and asking ourselves what sorts of lives we want to lead, or what sorts of persons we should be, then we can, over time, try to develop the kind of character and habits that we think best to achieve our goals. This means that on those (numerous) occasions when we do not stop and think about what to do, our conduct might nonetheless reflect a character we have either chosen to cultivate or allowed to develop. If we could not choose how to live and act on our decisions, then the habits and character traits that govern our conduct would not reflect our wills, and it would be hard to see how we could be responsible for them.

We believe it is realistic to assume that most adults are capable of important forms of self-governance. This is not because we believe that scientific findings are irrelevant to the truth of this assumption; we regard the claim that most adults are capable of self-governance as a straightforward factual claim, one that science might in principle show to be false. However, just as we need not be biologists to say that most humans are capable of walking or breathing, we need not be neuroscientists to be able to say, with authority, that most normal adult human beings can ask themselves what to do, make decisions, and act on their decisions, since most of us actually exercise these capacities on a regular basis. We can see that this is true through ordinary observation, without engaging in scientific inquiry.

What value, then, might conversation or collaboration between philosophers, neuroscientists, and those in related fields have for our understanding of self-governance? The answer is that neuroscientists, psychologists and others investigate many of the processes involved in self-governance, and learning from them about what the exercise of self-governance involves in the brain and the mind promises to illuminate and deepen our understanding of this capacity. Self-governance is not a capacity that we must either have or wholly lack. It can be challenged, diminished, even damaged in various ways, and when it is, we need to make difficult judgments about whether, and in what sense, to hold people responsible for what they do. Moreover, deepening our understanding of self-governance is crucial to knowing how to develop, strengthen, and repair weak or

[3] Greene and Cohen [23], p. 1781.

compromised self-governance, thereby enhancing our capacity for responsible agency.

Philosophers have long been interested in addictions, delusions, phobias, compulsions, and the like, as phenomena that point to compromised self-governance. Often, philosophers' accounts of compromised self-governance are, psychologically, fairly simplistic, describing (for instance) addicts or people with phobias as people who simply *cannot* choose to do certain things. What this inability comes to, and how, exactly, we are to conceive of and assess the deliberations, decisions, and actions of an agent who has it, are often not made clear. Yet it is not just in more dramatic cases – of the addict, or phobic, or delusional person – that self-governance is challenged and compromised. We often fall short in thoroughly ordinary circumstances: we say things we know we will regret later, feel angry when we know we have no reason to be, feel joy and glee in doing things we believe are wrong. Inner conflict is an ordinary part of being a human being. So is ambivalence. At times it can seem that we are "unable" to make a decision, or to follow through on what we decide; so too we can find ourselves "powerless" in the face of temptation, "drowning" in sorrow, "driven" by an obsession; stymied by anxiety, in ways that lead us to act against our better judgment, to forsake what matters most to us for what matters much less. Cognitive and volitional vulnerabilities can diminish self-governance, shaping what we attend to and how we construe what we attend to. Fear and aversion can, for example, lead to distorted interpretations of others as dangerous or threatening; listlessness and apathy may short-circuit forms of attentiveness crucial to empathy and compassion; resentment and aversion can undercut the fairness of our judgments. And if we value being clear-sighted, fair, empathic, and compassionate, we will, in such cases find ourselves making judgments and acting in ways we cannot reflectively endorse. In addition, our spontaneous reactions, our omissions and oversights, what we notice and neglect can sometimes be of great significance and yet are unchosen and sometimes unwitting [2]. Recent work in the cognitive and behavioral sciences provides insight into our inherent agential vulnerabilities revealing that we harbor and act on all kinds of implicit attitudes and biases that are often directly at odds with our explicit beliefs and avowed commitments [5, 6, 26–29, 30].

In some cases our failures of self-governance are episodic or occasional; in others they are longer-term. When we think of self-governance as an ongoing process, is it clear that it is itself often vulnerable to internal factors we experience as "out of our control."

When we exercise the capacity for self-governance, what we govern is *our own voluntary conduct*. It is not a failure of self-governance when a person has an epileptic seizure, or when she cannot control an involuntary reflex. This is so even though, in the case of the seizure, what causes her to behave as she does is her own brain. It is a failure of self-governance when a person succumbs to temptation against her own better judgment, or fails to carry out her own intentions when she could have done so. When this happens, we may feel alienated from ourselves, subject to "outlaw forces" [25]. But unlike when we are shoved on the elevator or slipped a drug that knocks us out, the "outlaw" forces that bear on self-governance are at least in important ways also internal to dimensions of our agency. They highlight ways we are vulnerable and limited in our capacities for self-knowledge and self-control, and suggest that self-governance is scalar, that it admits of degree and carries characteristic vulnerabilities and limitations. What impact such vulnerabilities should have on our moral responsibility is, of course, a contested normative question. But what is clear is that moral responsibility can be partial; the normative criteria we should invoke in deciding when it is mitigated or absent is a question that can be illuminated by deepened empirical study.

Thus the question arises: how might we distinguish cases in which we cannot govern ourselves from cases in which we fail to do so? When I say something "in spite of myself," fail to act when I believe I should, or am consistently disposed to impulsive behavior because of "internal" conditions that seem out of my control, how is this different from being subject to physiological states that produce reflexive responses or brain seizures? In what sense is my conduct voluntary in the former instances, but not in the latter? Neuroscientists cannot tell us what counts as voluntary behavior, or what counts as successful self-governance, because these are not scientific questions. But neuroscience can help identify parameters for distinguishing distinct ways in which we "lose control," whether we can regain it, and, if so, how.

The more we understand the neurological structures that underlie the capacity for self-governance, the more we will understand not just the nature of this capacity, but what its limits are, what its exercise requires, and the ways in which it can be impaired, restored, and strengthened. Understanding these issues will challenge and

clarify our understanding both of moral agency and moral responsibility, and of their limits. It will also help us to answer, more accurately and fairly, questions about the moral responsibility of persons whose capacity for self -governance is impaired. To the extent that we misdescribe such impairments, we are likely both to focus on the wrong theoretical questions about moral responsibility, and to give the wrong answers to moral questions about the responsibility of particular individuals on particular occasions.

Special Issue Contributions

The neuroscience community now has access to technologies and data of the necessary precision and magnitude to design studies with testable hypotheses about processes central to the capacity for self-governance: affective and behavioral self-regulation, integrated memory, attentional capacities, and a range of other facets of executive functioning and its relations to affective and cognitive functioning that have long interested philosophers. This special issue brings together a collection of authors and articles that take advantage of this science to advance our understanding of self-governance, but from different perspectives and different disciplinary backgrounds. Roskies [8] begins this special issue by addressing the worry that advances in neuroscience threaten to debunk our notion of ourselves as acting upon our own decisions—made after considered judgment of facts and values—in exercise of our will. She focuses not on the more commonly addressed threat of determinism to self-governance (and by extension free will), but rather on mechanism. Mechanism, Roskies explains, is distinct from determinism and is the view that "the mind/brain is some sort of machine or physical device, composed of interacting parts and governed by physical law." As molecular biology, genetics, and other fields have spent decades identifying and describing the mechanisms that generate beings, health, and disease, such a view seems justified. The question is whether mechanism undermines self-governance. Roskies argues that it does not. Key to the mechanism worry is an assumption that "mechanism entails mindlessness." Roskies uses the example of decision-making to show how progress in neuroscience might well suggest that mechanism is true "but that that mechanistic picture may be a small part of a larger mechanistic yet self-governing system, and that such systems are likely rich enough to undergird notions of agency and mindedness, and to support normative notions of responsibility." Roskies begins with a review of what the neuroscience of decision-making has taught us thus far and explains why this body of work may be taken to undermine agency. She then uses the metaphor of decentralized political systems to suggest how (the self and) self-governance may nonetheless be consistent with the mechanisms, processes, structures, and interactions among them that constitute decision-making. Roskies does this by reviewing a collection of such interacting processes and what we know about them from neuroscience. Roskies argues that these types of processes make self-governance possible, and further, that self-governance undergirds agency, and thus enables moral responsibility.

Sali et al. [31] continue the use of decision-making as an example for exploring self-governance, focusing on how biases in our attention and information gathering affect decision-making and thus our ability to behave in ways that align with our values (i.e., to be self-governing). Decision-making depends on noticing information in our environment and sorting through that information to collect and reflect on only that information which is relevant to the decision at hand. Determinations of relevance depend in part on our goals and moral beliefs. If we are not able to notice, make determinations of relevance, and weigh information in light of our goals and values, legitimate questions can be raised "about the degree to which the choices that we make may be poorly informed and not truly reflect our ability to otherwise exert self-governance." Sali et al. walk us through the data offered by neuroscience on attentional selection. As they note, "[a]ttentional selection sculpts our perception of the world around us, limiting what information reaches conscious awareness." What information reaches conscious awareness is influenced not only by the properties of the information/stimulus, but also by our goals and our ability to keep our goals in the front of our minds (in working memory). As most humans know, however, we do not always have perfect control over our attention. For example, a delicious donut or cookie in our present can make it difficult for us to keep in mind and prioritize our long-term health goals. While attentional control can be improved through the use of extrinsic rewards to reinforce attention "in a way that promotes self-governance when goals are consistent with reward associations," these same associations can work against

our best intentions when those associations conflict with current goals. Importantly, the degree to which reward history biases attention varies both across individuals and within individuals across the lifespan, suggesting that capacity for self-governance likewise varies. Sali et al. thus argue for an understanding of self-governance that takes into account and perhaps permits allowances for our individual histories and biases.

Niker et al. [32] shift our attention from our histories to our future selves, and from biases that can frustrate our attempts to behave in ways that align with our values to how the values themselves might be updated. The authors approach the challenge of changes to our values ("pro-attitude incorporation") and its importance for theories of autonomy from both philosophical and neuroscientific perspectives. They argue that self-governance requires "a self that is able to "update" itself in light of the world around it by responding to relevant experiences (as opposed to governance by an inflexible former self)" and further that conversation between philosophy and neuroscience is critical to ground philosophical theory in neuroscientific reality: the brain must be able to do what the philosophers are demanding of it. Niker et al. begin with the philosophical perspective, and an exploration of the role of pro-attitude revision and incorporation in theories of autonomy and our capacity for self-governance. In particular, the authors argue the importance of experience-responsive critical reflection as a feature of self-governance: we must be able to update our pro-attitudes in response to "relevant changes in the world around us." While current models for neuroscience research are not yet sophisticated enough to directly interrogate pro-attitude incorporation, Niker et al. summon evidence from experiments on information acquisition for perceptual and motor decisions to suggest how we may employ Baysian inference to incorporate new information in light of our prior beliefs about the world. The authors argue that current neuroscience offers a plausible mechanism for "incorporating new patterns of neural activity in response to novel sensory information." The relevant neuroscience data do, however, present a seeming challenge to the philosophers: whereas the philosophical literature tends to assume that pro-attitude incorporation requires "top-down, rational reflection", the neuroscientific data are consistent with the incorporation of new pro-attitudes "below the level of conscious awareness," similar to how we make routine (non-value-altering) decisions. Such challenges are critical to identify and grapple with if we are to achieve Niker et al.'s laudable call for productive conversation between philosophy and neuroscience and ultimately philosophical theory grounded in neuroscientific reality.

Fujita et al. [33] bring empirical evidence from psychology and related disciplines to bear on questions of self-governance, and self-control in particular. They argue that evidence to date does not support the much-discussed divided-mind (emotion versus cognition) combative model of self-control, and further that these data are better explained by a model characterized by structure and coordination, where consensus rather than conquest represents successful self-control. For example, a divided-mind model might posit that donut eating (giving into temptation) is driven by emotion, whereas diet maintenance (self-control) is driven by cognition; however, evidence suggests that being more thoughtful and deliberative can actually lead people to justify giving into temptation (a failure of self-control), rather than increase their chances of sticking to their long-term health-related goals. Fujita et al. argue that these data are better understood as a mutiny of parochial interests over the interests of the whole – a failure of structure, coordination, and consensus. Like Roskies, Fujita et al. employ a political metaphor to describe how their model functions, in their case to manage near-term temptations and long-term goals. They conjure a "senate-of-the-mind," where "[i]ndividual "senators" represent the various constituent elements of the mind, including people's wants and needs, thoughts, and behavioral tendencies," and "[p]olicy is driven by consensus." The senate-of-the-mind (and self-control) is successful when overall structure and cohesion is maintained. The authors outline the implications of their proposal for what constitutes a temptation and successful self-control, and how self-control under their model might be promoted or undermined. Fujita et al. employ social psychology's construal level theory (CLT), which posits that our decisions and behaviors are shaped largely by our subjective understanding of the world, in particular our psychological distance from events. High-level construal is associated with more distance and low-level construal with proximity. "Critically," Fujita et al. note, "CLT proposes that [a] change in construal may lead to changes in people's evaluations, judgments, and decisions." And indeed, data presented by Fujita et al. suggest that high-level versus low-level construal (i.e., maintaining a mindset that focuses on long-term versus short-term benefits) can, in fact, improve self-control in

a variety of ways. The authors present a case for a broader, perhaps more forgiving, view of self-control, marked by collaboration and give-and-take, rather than the conflict and winner-takes-all orientation of divided-mind models.

Like Fujita et al., Helion et al. [34] find the divided-mind, combative model of self-control insufficient. While they focus on the interplay between emotion and cognition, they argue for an emphasis on interaction over competition, give-and-take over winner-takes-all on the way to moral judgment. The authors begin by reviewing how others have viewed the relationship between emotion and cognition, the early history of which prioritized cognition in moral judgment and assigned emotion to a secondary role, generally in need of control. This balance then flipped, giving emotion the lead role, arguing that moral judgments are "made quickly and effortlessly and are the products of affective intuitions." Helion and Ochsner want instead to argue that emotion can be both automatic *and* controlled, and the questions are about how emotion emerges and to what degree it is controlled. Emotion, they argue can be both an automatic response (fear of the spider on the bed) or the result of cognitive processes (fear induced on sleepless nights by one's probability calculations about the presence of spiders). Further, they argue that control is bidirectional: cognitive processes shape and change affect and affect shapes and changes cognitive processes. How this shaping proceeds depends in part on what other authors in the special issue would call our pro-attitudes and reward histories, and is supported by multiple regions of the brain. And like other aspects of self-governance, this process of shaping and give-and-take varies both across individuals and within individuals throughout their lives, both developmentally and in response to the different roles we assume over time. Helion and Ochsner argue that their model for thinking about the role of emotion in moral judgment can be used to motivate future research in psychology and neuroscience that will generate new insights into this uniquely human behavior.

Implications

The goal of this special issue and the larger area of inquiry we describe is to understand the bearing of advances in neuroscience and related fields on our capacity for self-governance, and to use that knowledge to develop a more complete understanding of how that capacity is developed, and how it can be damaged, restored, and sustained. We expect that such an understanding would have valuable applications in a variety of fields.

First, it is useful to moral philosophy generally. Moral philosophers seek to understand how we ought to exercise our capacity for self-governance. As we have noted, understanding what advances in neuroscience show us about the nature and limits of that capacity, and about the ways in which it can be supported, challenged and diminished, would be of great value in helping us understand how to encourage the development and exercise of this capacity, and what we should conclude, morally, in cases of failure. Moreover, a clearer understanding of the conditions under which someone may justly be held responsible for her character and conduct would be useful to philosophers working on such topics as guilt, shame, blame and forgiveness. Perhaps one lesson from this special issue to which philosophers should attend is that many of the capacities required for self-governance vary across individuals and within individuals over time and in response to experience.

Second, clarity on these questions would have implications for those areas of public policy in which it is important to distinguish conduct for which people can legitimately be held morally responsible, from behavior that is, in some sense, out of their control. As advances in neuroscience can help us understand the nature and limits of our capacity to govern our own behavior, work of the sort included here will, we hope, will be useful to scholars who seek to clarify both the kinds of behavior policy makers and analysts might legitimately expect people to change and strategies that would be effective in helping them – for example, in public health interventions to decrease rates of obesity or heart disease, in efforts to rehabilitate convicted criminals, treat PTSD, or address domestic violence or substance abuse. Of note, work in this special issue suggests that our model of decision-making may have significant implications for which circumstances and efforts can strengthen or diminish our capacity for self-governance.

Third, this work can be useful to those neuroscientists who are interested in bringing their discoveries to bear on moral questions. Just as philosophers' treatments of self-governance and moral responsibility have been hindered by their unfamiliarity with advances in neuroscience, neuroscientists who write about the relation

between their work and ethics are often unfamiliar with the philosophical literature on the topics they address, and with the arguments for and against various philosophical views. We hope that providing contributing to an account of self-governance that is both neurologically and philosophically sophisticated might make those arguments more accessible to neuroscientists.

In aiming to understand more comprehensively and precisely the ways in which neuroscience sheds light on the workings of those processes constitutive of self-governance, the papers in this special issue help to illuminate and clarify ways in which science can deepen our conception of ourselves as moral agents, and inform effective approaches to developing the capacity for self-governance, thereby *strengthening*, rather than weakening, both our understanding of ourselves as morally responsible agents and our strategies for fostering the realization of morally responsible agency.

Acknowledgements We wish to express our gratitude to the National Endowment for the Humanities and the Henry R. Luce Foundation (through a Professorship to HB) for their generous support for this work. Special thanks to editor Neil Levy for his helpful comments on an earlier version of this paper.

References

1. Scanlon, T. 1998. *What we owe to each other*. Cambridge: Harvard University Press.
2. Smith, A. 2005. Responsibility for attitudes: Activity and passivity in mental life. *Ethics* 115 (2): 236–271.
3. Smith, A. 2012. Attributability, answerability, and accountability: In defense of a unified account. *Ethics* 122: 575–589.
4. Smith, A. 2015. Attitudes, tracing, and control. *Journal of Applied Philosophy* 32 (2): 115–132.
5. Levy, N. 2015. Neither fish nor fowl: Implicit attitudes as patchy endorsements. *Nous* 49 (4): 800–823.
6. Levy, N. 2017. Am I a racist? Implicit bias and the ascription of racism. *The Philosophical Quarterly* 67 (268): 534–551.
7. Levy, N. 2005. The good, the bad and the blameworthy. *Journal of Ethics and Social Philosophy* 1 (2): 2–16.
8. Roskies, A. 2016. Decision-making and self-governing systems. https://doi.org/10.1007/s12152-016-9280-9.
9. Fischer, J.M., and M. Ravizza. 1998. *Responsibility and control*. Cambridge: Cambridge University Press.
10. Watson, G., ed. 1982. *Free will*. Oxford: Oxford University Press.
11. Honderich, T. 1988. *A theory of determinism*. Oxford: Oxford University Press.
12. Pereboom, D. 2001. *Living without free will*. Cambridge: Cambridge University Press.
13. Kane, R., ed. 2011. *The Oxford handbook on free will*. Oxford: Oxford University Press.
14. Kane, R. 1998. *The significance of free will*. Oxford: Oxford University Press.
15. O'Connor, T. 2000. *Persons and causes: The metaphysics of free will*. Oxford: Oxford University Press.
16. Bok, H. 1998. *Freedom and responsibility*. Princeton: Princeton University Press.
17. Strawson, G. 1986. *Freedom and belief*. Oxford: Oxford University Press.
18. Smilansky, S. 2000. *Free will and illusion*. Oxford: Oxford University Press.
19. Roskies, A. 2004. Everyday neuromorality. *Cerebrum* 6 (4): 58–65.
20. Roskies, A. 2002. Neuroethics for the new millenium. *Neuron* 35 (1): 21–23.
21. Bourget, D., and D.J. Chalmers. 2014. What do philosophers believe? *Philosophical Studies* 170 (3): 465–500.
22. Roskies, A. 2006. Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Science* 10 (9): 419–423.
23. Greene, J., and J.D. Cohen. 2004. For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society B* 359: 1775–1785.
24. Nichols, S. 2006. Folk intuitions on free will. *Journal of Cognition and Culture* 6: 57–86.
25. Frankfurt, H. 1971. Freedom of the will and the concept of a person. *The Journal of Philosophy* 68 (1): 5–20.
26. Brownstein, M., and J. Saul, eds. 2016b. *Implicit bias and philosophy: Moral responsibility, structural injustice, and ethics*. Oxford: Oxford University Press.
27. Dasgupta, N. 2013. Implicit attitudes and beliefs adapt to situations: A decade of research on the malleability of implicit prejudice, stereotypes, and the self-concept. *Advances in Experimental Social Psychology* 47: 233–279.
28. Huebner, B. 2016. Implicit bias, reinforcement learning, and scaffolded moral cognition. In *Implicit Bias and Philosophy: Metaphysics and epistemology*, ed. M. Brownstein and J. Saul, 47–79. Oxford: Oxford University Press.
29. Mandelbaum, E. 2016. Attitude, inference, association: On the propositional structure of implicit bias. *Nous* 50 (3): 629–658.
30. Brownstein, M., and J. Saul, eds. 2016a. *Implicit bias and philosophy: Metaphysics and epistemology*. Oxford: Oxford University Press.
31. Sali, A., Anderson, B., and Courtney, S. 2016. Information processing biases in the brain: Implications for decision-making and self-governance. https://doi.org/10.1007/s12152-016-9251-1.
32. Niker, F., Reiner, P., and Felsen, G. 2016. Updating our selves: Synthesizing philosophical and neurobiological perspectives on incorporating new information into our worldview. (this volume). https://doi.org/10.1007/s12152-015-9246-3
33. Kentaro Fujita, Jessica Carnevale, and Yaacov Trope. 2016. Understanding Self-Control as a Whole vs. Part Dynamic. https://doi.org/10.1007/s12152-016-9250-2.
34. Helion, C. and Ochsner, K. 2016. The Role of Emotion Regulation in Moral Judgment. https://doi.org/10.1007/s12152-016-9261-z.