The Philosophical Quarterly Vol. 0, No. 0
ISSN 0031-8094 doi: 10.1093/pq/pqw048

THE CAUSAL MAP AND MORAL PSYCHOLOGY

By Timothy Schroeder

Some philosophers hold that the neuroscience of action is, in practice or in principle, incapable of touching debates in action theory and moral psychology. The role of desires in action, the existence of basic actions, and the like are topics that (they hold) must be sorted out by philosophers alone: at least at present, and perhaps by the very nature of the questions. This paper examines both philosophical and empirical arguments against the relevance of neuroscience to such questions and argues that neither succeeds. In practice, there is already a stable body of findings from neuroanatomy and neurophysiology that warrants attention. And as a matter of principle, the 'causal map' of action production derivable from these findings requires the study of action theorists and moral psychologists because every such philosopher has commitments (sometimes, deeply implicit) to the shape of this causal map: commitments that might be in conflict with reality.

Keywords: action, neuroscience, Wittgenstein, Davidson, functionalism, causation.

The neuroscience of movement has made an enormous amount of progress over the last fifty years, at every level of investigation. What does this signify for philosophy? A few philosophers think that we should read our philosophy of action and moral psychology more or less directly off of the neuroscience, abandoning common sense categories when they do not fit neatly with what neuroscience reveals.[1] I will assume for this paper that this radical position is false, though of course there is much more to say about it.[2] More philosophers think that specific findings in the neuroscience of action have important implications for the philosophy of action and moral psychology, while holding that this is a matter of philosophy and neuroscience meeting on more or less equal terms.[3] And then there is a large group of philosophers holding that there is little or nothing of interest to the philosopher of action or moral psychologist in neuroscience.

[1] This is approximately the position of Churchland (1986, 2002) and Churchland (1995, 2007).
[2] One powerful but radical objection to Churchland-style eliminative materialism comes from Stich (1996). A less radical version of related ideas is found in Schroeder (2004).
[3] E.g., Greene (2010), Holton (2009), Roskies (2003), Railton (2014), Schroeder (2004), and Walter (2001).

Christine Korsgaard draws this view particularly vividly. In *Sources of Normativity* (1996) she writes:

> The freedom discovered in reflection is not a theoretical property which can also be seen by scientists considering the agent's deliberations third-personally and from outside. It is from within the deliberative perspective that we see our desires as providing suggestions which we may take or leave. You will say that this means that our freedom is not 'real' only if you have defined the 'real' as what can be identified by scientists looking at things third-personally and from outside.

Korsgaard is hardly alone. According to Hornsby (1997), it is ultimately wrong to think that 'the basis of our everyday understanding of one another is susceptible to correction and refinement by experts in some specialist field where empirical considerations of some non-common sense kind can be brought to bear.' And, as will be seen, action theorists and moral psychologists such as Donald Davidson, Peter Hacker, and Philip Pettit have made related claims.

In this paper, I argue that this third position is mistaken. I consider reasons philosophers might have for holding that they need not attend to neuroscience in the philosophy of action or moral psychology, dividing them into two groups. Some of these reasons involve purely philosophical considerations; they are poor reasons, on balance, because of philosophical considerations that have been overlooked or neglected. Others of these reasons involve the empirical facts; they are poor reasons, on balance, because of how the empirical facts actually stand.

I. THE RELEVANT NEUROSCIENCE

The neuroscience that I wish to argue is relevant to action theorists and moral psychologists is neuroscience that maps out, at the level of individual neurons and small groups of neurons, the flow of cause and effect leading to the movement of the body. This is the domain of the low-level neuroanatomy and neurophysiology of movement. Neuroanatomy is the study of the individual parts of the nervous system, and neurophysiology the study of the functional interrelations among the parts. Together, these branches of neuroscience, when applied to the study of movement, aim to provide what amounts to a comprehensive map of the neurons that influence movement (neuroanatomy), annotated to show how each part of the map contributes to the causation of movement (neurophysiology). Call this sought-for annotated map 'the causal map'.

Notice that the causal map would be a more powerful tool than many others. Unlike simple brain-imaging studies, for example, the causal map

would not traffic in mere correlations.[4] The causal map would show direct causal relationships, relationships that are as observable as anything in the natural sciences.

Naively, one might think that action theorists and moral psychologists would be eager for the causal map. A number of questions in action theory and moral psychology would appear best answered with its help. Are there basic or primitive actions? Are prior intentions motivating on their own? Do all actions require (true) desires to be undertaken? Are there three fundamental sources of bodily movement, aptly called 'appetite', 'passion', and 'reason'? Is there something that distinguishes the person who has reasons A and B to perform an action and performs it for both from the person who has the same reasons but performs the action for only one? These appear to be questions that one would answer best by drawing upon facts about cause and effect on the way to movement, in conjunction with purely philosophical considerations. And facts about cause and effect on the way to movement are exactly what a philosopher might learn from studying the causal map.

Of course, the causal map could not provide answers to the above problems on its own. The labels that low-level neuroanatomy will place on the causal map will be things such as 'central nucleus of the amygdala', not 'the passions'. But, the naive philosopher might reason, the causal map would still provide a helpful and needed set of facts.

For a concrete example, consider efforts by philosophers to interpret the brain's release of dopamine. The release of dopamine is carried out by a small cluster of cells deep in the brain, but appears to have a broad range of effects on action, feelings of pleasure, the direction of attention, addiction, and more.[5] Most strikingly, dopamine release is a normal requirement for paradigmatic voluntary action. This essential contribution of the dopamine system has suggested a number of psychological interpretations of the system (or of what it realizes, or what supervenes on it, or. . .). Philosophers have interpreted the activity of the dopamine system as, for instance, episodes of pleasure,[6] consistent with a hedonistic theory of motivation, as expressive of changes in net apparent satisfaction of intrinsic desires,[7] consistent with a broadly neo-Humean theory of motivation, and as judgements of what it is best to do,[8] consistent with a broadly neo-Socratic theory of motivation.[9] Considering

[4] A weakness pointed out in Berker (2009) among many other works.

[5] Morillo (1990) offers the first philosophical overview of the relevant science, and Arpaly and Schroeder (2014: ch. 6) offers a recent philosophical overview aimed at moral psychologists. An excellent scientific overview can be found in Schultz (2015).

[6] Morillo (1990).

[7] Schroeder (2004).

[8] Yaffe (2013).

[9] Other valuable attempts by philosophers to interpret the dopamine system include those of Butler (1992), Dill and Holton (2014), Holton (2009), Levy (2014), and Shea (2014).

only the effects of the reward system upon action, each theory has initial credibility. However, with the full causal map in hand, the naive philosopher might hope that some of these rival views would gain support, while others would encounter obstacles, thus clarifying if not resolving the debate. Perhaps, the full causal map would reveal that good candidates for the neural bases of judgements about practical reasons or the pro tanto goodness of particular courses of action, but not the neural bases of sweet taste experiences, tend to directly cause activations of the dopamine system. That would, to the naive philosopher, appear strong evidence in favour of a neo-Socratic interpretation of such activity given common philosophical commitments about the causal network in which beliefs about what actions are best are likely to be found. Or perhaps the full causal map would reveal that the neural bases of sweet taste experiences, massage-like tactile experiences, neural representations of restorations of blood sugar and basic hydration, and so on, are by far the predominant direct causes of activation of the dopamine system. That would appear some evidence for the hedonistic interpretation of its activity. And so on. Perhaps, the naive philosopher would concede, the causal map will prove quite ambiguous, or its philosophical interpreters will be sufficiently ingenious to overcome apparent empirical problems. But even so, the causal map is the sort of thing that a philosopher would turn to as a check against unconstrained theorizing. As of now, the philosophers interpreting the dopamine system have not focused on comparing and contrasting how their interpretations fit better or worse with the larger causal map, but the naive philosopher might well hope that such comparisons are coming and will be enlightening.

This naive picture of how philosophers might approach the prospect of the causal map is quite naive. While some philosophers (such as those debating the interpretation of the dopamine system) have embraced this partly empirical methodology, many others have not. These other philosophers have found many reasons to doubt that they have any use for the causal map. Answering these doubts is the purpose of this paper.

II. PURELY PHILOSOPHICAL REASONS FOR PHILOSOPHERS TO IGNORE THE NEUROSCIENCE OF MOVEMENT

Empirical objections will come later. For now, consider purely philosophical objections to the relevance of the causal map. The main such objections I want to address are those stemming from three broad positions in the philosophy of mind. In order of their rising to prominence, these positions are the Wittgensteinian family of theories, interpretationism, and strongly a priori functionalism.

Before beginning, just a word about what is going on: the claim is not that the action theorists and moral psychologists with whom this paper is arguing

universally commit themselves to these theories of mind. There are some who commit themselves to particular theories but many more who do not. The claim, rather, is that these three approaches to the mind provide the three best (and most prominent) apparent justifications for disregarding the causal map. Thinking about how these theories of mind ultimately fail to insulate philosophy from the causal map is thus the most efficient way to demonstrate its relevance.

So, first consider the Wittgensteinian family of theories of mind. For my purposes, this includes philosophical behaviourism of Ryle's (1949) sort, the positions of Wittgenstein (1953)[10] and many contemporary Wittgensteinians (e.g., Bennett and Hacker 2003), and some descendants of the Wittgensteinian research program (e.g., Brandom 1994). Of course, there are enormous differences between these various views. But they can be grouped together here because they afford two shared arguments that the causal map would be irrelevant to philosophers of action and moral psychologists.

The first argument derives from a strongly a priori approach to the nature of the mind that privileges behaviour (actual and possible), along with the external context of that behaviour (especially, the linguistic, social, and normative context), in making claims about the mind. Bennett and Hacker express a widely shared thought within this approach when they write, 'The primary grounds or evidence for the ascription of psychological predicates to another are behavioural' and this behavioural evidence is criterial, that is, 'logically good evidence' (Bennett and Hacker 2003). Since behaviour entails nothing about the underlying neuroscience beyond the bare fact that it permits the behaviour in question, it has seemed to many in the Wittgensteinian family that there can be no good argumentative route from neuroscientific facts back to claims about the mind that might challenge philosophers. Bennett and Hacker (2003) again put the point succinctly. If there would be strong behavioural evidence for a psychological predication but strong neuroscientific evidence against the psychological predication, 'the latter is defeated' and the inductive correlations on the basis of which the latter came to see convincing 'need to be re-examined'.

The second argument from a broadly Wittgensteinian perspective invokes the idea that mental states in general are something like dispositions (Ryle), abilities (Bennett and Hacker), or skills (Brandom), and so are not even in the right ontological category to have counterparts on a map of neurons. Likewise with the idea that mental events are exercises of these same dispositions, abilities, or skills in a physical, linguistic, social, or normative context that is constitutively essential to their identities as the mental events they are. Such

[10] Though, as the reader will see, I will not make an effort here to interpret *Philosophical Investigations*. For the purposes of this paper, I take Peter Hacker's interpretation of that and related works to be an adequate proxy.

ideas about the mind would seem to remove mental explanations from the domain of neural causation so completely that the irrelevance of the causal map is guaranteed.

Now consider the second broad family: positions that are, or are close to, Donald Davidson's interpretationism (e.g., Davidson 1980; Davidson 1984; Dennett 1987; Hornsby 1997;[11] Mölder 2010). On Davidson's own formulation of the view, every token mental event (a particular choosing to type the word 'Donald', for instance) is identical to a token neural event (one that is aptly poised for causing the finger movements involved in typing 'Donald', for instance). The thesis of token-identity might appear to guarantee the relevance of the causal map to theorizing about action, but of course this does not obviously follow. After all, philosophical theorizing concerns, not tokens, but types. And, according to the interpretationist, whether one has a certain type of mental event is settled, not by the neural facts, but by interpretive principles that appeal to ordinary practices of understanding each other (as reasonable, lovers of the good, believers of the true, and so on). As Davidson (1980) famously writes,

> There are no strict psychophysical laws because of the disparate commitments of the mental and physical schemes. It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and conditions physically described. It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs, and intentions of the individual.

Facts about what it would be reasonable to be thinking and wanting while moving thusly in a context are the sorts of facts that constitute what mental events one undergoes when so moving. This might mean that, on one occasion, a certain type of neural event is token-identical to a belief that it would be best to do A, and, on another occasion, that same type of neural event would be token-identical to a mere desire to do A. And so it seems the interpretationist is, in fact, insulated from the relevance of the causal map. It can be reinterpreted as she needs to fit the interpretive demands of each new situation.

The third view to consider is the variety of causal-role functionalism that gives a high priority to claims said to be known a priori when theorizing about the mental states of interest to action theory or moral psychology. Frank Jackson and Philip Pettit exemplify this approach in 'Moral Functionalism and Moral Motivation', for instance, suggesting that concepts such as that of fairness are constituted by being at the centre of a causal-functional network that includes both intellectual and motivational features. To truly possess the concept one must judge (in an unthinking, not highly intellectualized manner)

[11] Hornsby is an interpretationist who rejects Davidson's idea that there are token-identities between mental and neural (or physical) events. Thus, for some purposes Hornsby will prove closer to the broadly Wittgensteinian theorist than to Davidson in what follows. This complication should not change the overall force of the arguments to come, however.

things to be fair on the basis of their having or lacking certain features, and also be motivated (though not necessarily all the way to action) to bring about what one judges to be fair and shun what one does not. They hold that they can stipulate this because, as causal-role functionalists, they hold it is true of all mental states that the states get their contents and their attitudinal roles from their causal-functional roles. On their picture, then, 'there is no way of judging non-intellectually that something is fair without experiencing a suitable desire for the option in question. The idea of forming a fairness-belief in this non-intellectual way, and yet lacking the desire, will be . . . incoherent' (2004). Thus, philosophers figure out the truths of action theory and moral psychology, and then neuroscientists are permitted to take the functional roles on which the philosophers have settled and find their realizers in the brain—realizers that we can be assured exist, since something (if only a rather neurologically gerrymandered something) must underlie the behavioural, emotional, and intellectual patterns that led philosophers to their functional theories in the first place.

Two remarks should be made about these three positions, just to keep things clear.

The first is that, although certain theories of mind are famously associated with certain theories of action or claims in moral psychology, there is nonetheless a great deal of independence between them. Davidson has familiar positions on primitive actions and weakness of the will, for instance, but these are not commitments of interpretationism as such. Almost all of the usual debates in action theory and moral psychology are live debates under all three of the above theories of mind.[12] The relevance of the above theories is not that they settle what one should hold as an action theorist or moral psychologist, but rather that they represent the best strategies available to insulate the debates internal to action theory and moral psychology from the facts provided by the causal map, rendering it irrelevant.

The second is that the goal of this paper is to be as inclusive as possible: to show that action theorists and moral psychologists of all stripes need to consider the causal map. Hence, it would be illegitimate to argue for the relevance of the causal map by, e.g., arguing that there is something flawed in Ryle's behaviourism.

III. CAUSAL COMMITMENTS

It appears, then, that there are multiple systematic theories in the philosophy of mind on which the causal map is irrelevant to action theory and moral psychology.

[12] The obvious exception is debates about whether a theory of action should be causal.

To begin to challenge this appearance, consider an example. Suppose that Farhad consciously thinks it best to raise his hand now, and that just after so thinking he raises his hand, and so votes to go on strike (that being why he thought it best to then raise his hand). About this example, some philosophers would say that the whole story has been told. Farhad thought a particular course of action was best, and that alone was enough (perhaps, in a broadly rational person) to ensure that he raised his hand. Other philosophers would argue that there is a key missing piece: an intrinsic desire to do what is best, one that pre-exists Farhad's thinking it best to raise his hand. There are, famously, a number of lines of purely philosophical argument favouring the one position or the other, and these arguments are entirely neutral between broadly Wittgensteinian, interpretationist, and functionalist ways of thinking about the mind. Is there a way that the facts depicted on the causal map might be found to matter to these theorists, appearances notwithstanding?

Consciously thinking it is best to now raise one's hand—or anything else—has certain features. Conscious thought is typically remembered, at least for a little while, as an episode in one's life. It would surprise no one if Farhad could remember what he thought two hours later. Thinking something consciously also leads to associated thoughts. For instance, it might have occurred to Farhad to wonder whether his nails were dirty, exactly because he thought about raising his hand. Conscious thinking is particularly well poised to stir up memories. Perhaps, Farhad's consciously thinking it best to raise his hand and so go on strike leads him to remember the last time he was on strike, or the last time he voted by show of hands. And conscious thinking is particularly well poised to influence one's emotions. It would be no surprise if Farhad's thought made him anxious, for instance.

Holding these to be features of some conscious thoughts commits a causal-role functionalist to holding that the token neural state or event that realized Farhad's conscious thought was, on that occasion, positioned in a causal network in a way consistent with the thought having all of these features. Likewise, it commits interpretationists[13] to holding that the token neural state that was token-identical to Farhad's conscious thought was, on that occasion, positioned in a causal network in a way consistent with the thought having all of these features. These very modest commitments to neural-causal facts follow directly from these sorts of theories, since they both embrace the idea that at least token mental events have ordinary causal roles.

Less obviously, philosophers of a broadly Wittgensteinian stripe also have modest causal commitments to neural-causal facts holding in Farhad's case (and so, in every case).

[13] Again, setting aside Hornsby (1997) here and Dennett (1987). For this purpose (only), their positions can be grouped with that of the Wittgensteinian family.

The philosopher who follows Ryle in holding that mental states are dispositions to act will hold, following the second chapter of Ryle's *Concept of Mind* (1949), that these are specific, complex dispositions to act. And at any moment, a given complex disposition of the required sort will have some ground, found at least partly inside the skull of the person with the disposition.[14] This ground of the relevant dispositions in Farhad will include some neural state, embedded in a network of causal relationships. Thus, a commitment to a mental state entails a commitment to there existing some neural state suited to make true a specific set of claims about dispositions. This neural state has to have causal features suited to make these claims about dispositions true. Thus, commitment to it is commitment to some claims about causes, however modest.

Philosophers following Bennett and Hacker hold that a person who raises his hand intentionally has to be exercising an ability to raise his hand, or else is merely subject to a tic or something similar (2003). Similarly, Robert Brandom's *Making it Explicit* uses the idea of 'reliable differential responsive skills' (1994). But now, consider what gives a person such abilities or skills. The grounds of these abilities or skills include a great deal of what lies in the brain. They also include a great deal of what does not lie inside the brain, according to these theorists: the existence of language games and forms of life, or a community that creates genuine normative statuses for the exercise of the abilities or skills, or similar things. Still, having an ability or skill requires the existence of neural states that might ground such an ability or skill's most local manifestations. And so, holding, say, that a person raised his hand purely out of a sense of duty, and not at all because of what he wanted, is holding that one set of abilities or skills, and not a different set of abilities or skills, was used—and so the narrow neural ground of one specific set of abilities or skills was causally involved in the hand's raising, and not the ground of a different specific set of abilities or skills.

Now consider just one specific detail from Farhad's story in the light of the above weak causal commitments. Imagine that Farhad remembers his conscious thought that it would be best to raise his hand: he can later recall thinking just that thought. A conscious thought that is remembered has to cause (or be realized by, grounded by, etc., something that causes) the formation of a memory (or the realizer of, or ground of, etc., a memory). As it happens, episodic memory in ordinary human beings relies on neural changes in the hippocampus and adjacent regions. Without the generation of these changes, no new episodic memories will be formed (Kandel *et al.* 2013). These changes are part of the narrow realizer, ground, or etc. of recently formed episodic memories. Thus, a remembered conscious thought always in fact causes (or is

[14] Perhaps there are ungrounded dispositions in basic physics or the like. But there are no ungrounded dispositions to raise one's arm when a vote is called.

realized by, grounded in, etc., something that causes) appropriate changes in the hippocampus.

The above sort of argument can be repeated for each of the ordinary features attributed to conscious thoughts. This generates a network of apparent causal commitments, just given commitment to the claims that Farhad thought to himself that it would be best to now raise his hand, that he remembered his thought, that the thought made him wonder whether his fingernails might not be dirty, that his thought reminded him of the last time he voted, and that his thought made him feel a stab of anxiety. The neuroscience of the specific neural preconditions of these various mental processes is not, at present, as clear as the neuroscience of the preconditions for episodic memory. But it will be, eventually, just as clear. And it is already clear that, for example, if the stab of anxiety involves feeling a (real) sudden tightening of the stomach, that feeling follows non-coincidentally from Farhad's thought only if Farhad's thought causes (or is realized by or grounded in etc. something that causes) changes in the amygdala that in turn cause changes in blood flow around the stomach that in turn cause perceptions of these changes (or neural states that realize or ground perceptions of these changes).[15]

Now consider the contrastive claim that Farhad's raising his hand was because of his judgement that that was best, and *not* dependent on a pre-existing standing desire to do what is best.

Any claim about a pre-existing standing desire to do what is best, like any claim about a conscious thought, will come with its own causal commitments. For the sake of example, imagine two popular ones. For something to be a desire, as opposed to a belief about goodness, it must be the sort of thing that disposes one to pleasure, should one get what is desired. And for something to be a desire it must have an influence upon action that is structurally similar to the influence had by the states that move one to eat unpalatable food when hungry, or to drink unpalatable liquids when thirsty (two canonical desires).[16]

From these philosophical claims come more causal commitments. For Farhad to lack an intrinsic desire for what is best, or for him to have such a desire without it playing any part in his having raised his hand, it is necessary that there not have been a neural structure, distinct from that which is, realizes, or grounds etc. his thought that raising his hand is best, possessing the causal properties committed to by the working assumptions about standing desires. Any other neural state or event playing an important causal role in getting Farhad's hand to raise must lack certain features: not be closely connected to the causation of neural activity that is, realizes, or grounds (etc.) pleasure, or, if it is so connected to pleasure, not also have causal relations structurally

[15] LeDoux (1996) provides one accessible presentation of the neuroscientific details.

[16] Here I draw partly on works such as Davis (1986) and Schueler (1995) where much is made of what distinguishes desires proper from more cognitive states that might also motivate.

parallel to those had by the neural structures that are the best candidates for the neural realization, or grounds (etc.) of hunger or thirst.

These causal commitments, and those found just in committing to the existence of a thought that some course of action is best, are not trivially satisfied. They might jointly be satisfied in a given person, like Farhad, at a given time, such as at the moment when Farhad votes, or they might not. If they are not satisfied, there might be adjustments to make to the working theories of conscious thought or standing desire that would not seem outré and that would generate causal commitments that are all satisfied. In that case, the action theorist or moral psychologist might have learned that she needs a slightly interesting theory of thoughts or standing desires as a part of her larger theory. Or it might be that no adjustment that leaves intact the idea of ordinary conscious thoughts and of ordinary desires can be held consistently with holding that Farhad raised his hand independently of any standing desire. Which way things will turn out will depend in part on what claims about thoughts and desires are acceptable to the philosopher, and in part on the causal map.

In short, much of action theory and moral psychology has inescapable causal commitments. And because these causal commitments can only be seen to be consistent or inconsistent with all the facts when one knows the details of the web of causal relationships available to be, realize, or ground the relevant philosophical claims, the consistency or inconsistency of many philosophical doctrines depends upon the facts provided by the causal map.

IV. OBJECTION: GERRYMANDERING

It might appear that the foregoing argument has presupposed that the scientifically privileged anatomical and physiological features of the brain will always be (realize, ground, etc.) the truth-makers for the claims philosophers want to make. If this was in fact presupposed, though, it was presupposed illegitimately. So far as the behaviourist, interpretationist, or functionalist is concerned, there is no reason to privilege anything considered natural or unified at the neural level over the needs of psychological-level interpretation. A neuroscientist might balk at treating a state of the hippocampus and a state of the amygdala as just one neural state, but a philosopher of any of the above stripes need have no objection to forcing the unification of these states into a single mental state, realizer of a mental state, or narrow, partial ground of the skills constituting a mental state, if necessary for philosophical purposes.

An example will help to make things slightly more concrete. Suppose one philosopher holds that Farhad acted solely on his thought that his action was best, while another philosopher holds that Farhad acted in part on the thought and in part on a standing desire to do what is best. And suppose

that the second philosopher says to the first, 'You claim that thought alone produced this action, while other actions are influenced by desires as well. So saying commits you, given the causal map, to saying that neural structures A and B played a role, while no neural structure playing roles X and Y also played an important contributory role in causing the raising of the hand. However, the causal map reveals that one neural structure playing roles X and Y did play a role in Farhad's raising his hand. Thus your interpretation of his action is mistaken, and mine is vindicated'. Here, the first philosopher might say something like, 'neural structures X and Y played no *distinctive* causal role, because I interpret one or both of them as, in this particular instance, being a part of (or part of the realizer of, or part of the ground of, etc.) the attitude that I claim led to the action on its own, namely, the thought'.

In general, if it appears that a philosophical claim problematically commits a philosopher to ignoring causally important neural structure X, then a defender of the claim can simply hold that structure X is, at least on this occasion, part of (etc.) the thing the philosopher was committed to saying was causally important—and so X is not being ignored after all. And likewise, if it appears that a philosophical claim commits a philosopher to giving a prominent role to a causally unimportant neural structure, W, then a defender of the claim can simply hold that W is a part of a larger neural structure, Z, and Z is not causally unimportant. Similar manoeuvres can be deployed by the philosopher defending claims about the unity or multiplicity of types of causes (or realizers, grounds, etc. of causes) of movements, and so on.

An overly sanguine attitude toward this sort of gerrymandering will not do, however. While in principle there is nothing wrong with holding that there is a loose fit between neural types and psychological types (only a very strict type-identity theory holds otherwise), gerrymandering will not fix problems created by the gerrymandering philosopher's own philosophical commitments.

In the case of theorizing about Farhad's raising of his hand, the philosopher who holds that it was done only on the basis of the thought that it was best, and not also from a standing desire to do what is best, has some account, perhaps sketchy and tentative in nature, of what a thought is, what a standing desire is, and what their main differentiating features might be. And it was from these perhaps sketchy and tentative commitments, joined with the facts set out in the causal map, that an inconsistency was generated by the critic. (At least, this was how it was imagined to go. Perhaps, it is the critic who would be found to be inconsistent, after a closer inspection of the causal map, or both or neither.) Thus, proposing to gerrymander things so that neural structures X and Y are counted as a part of the thought is proposing to gerrymander things so that the thought turns out to have, on this occasion, both the features attributed to thoughts but not standing desires and the features attributed to standing desires but not thoughts. This is clearly unacceptable.

When Baker and Hacker argue that ordinary evidence for psychological claims must necessarily defeat inconsistent claims made by neuroscience, they implicitly presuppose that there is some way of holding all of the psychological claims they think are correct simultaneously. But this is not guaranteed. It all depends on what the causal commitments of these claims might be, and whether these commitments can or cannot be satisfied. Likewise, when Jackson and Pettit hold that it would be incoherent to attribute a thought with the concept of fairness except under certain empirical conditions, they implicitly presuppose that attributing a thought involving the concept of fairness under those conditions can be done consistently with the other things they think are true of such thoughts and of minds in general. But again this is not guaranteed. Again, it depends on what their causal commitments might be, and whether those commitments can be satisfied. It depends on the commitments the philosophers hold, and on the facts represented by the causal map.

Thus, gerrymandering does not, and cannot, solve the problems that I have been suggesting are generated by considering the causal map because those problems do not arise independently of a philosophical theory about what thoughts, standing desires, emotions, immediate intentions to act, intentions to act in the future, feeling of pleasure, and so on are like. It is only in conjunction with philosophical commitments about what these mental states and events are like that the causal map can generate inconsistencies in a philosophical position. Given these commitments, however, gerrymandering the interpretation of the neuroscience so as to remove the inconsistencies inevitably amounts to gerrymandering the commitments to the natures of the various relevant mental states and events. And thus, gerrymandering that is motivated just by the wish to remove these sorts of inconsistencies can move the location of, but not change the fact of, internal inconsistency in a view.

V. OBJECTION: TYPES AND TOKENS

Another way of objecting to the line of argument in Section III is to hold that it illegitimately shifted between something like token-identity (or -realization, or -grounding) to something like type-identity. And since no behaviourist, interpretationist, or a priori functionalist is committed to type-identities (or type-realizations, or type-groundings), this illegitimate shift vitiates the argument.

It certainly does appear that there was an illegitimate shift of this sort in the discussion of Farhad's raising of his hand. A fact about types was brought into the discussion in order to generate the possibility of conflicting commitments in Farhad's token case. Specifically, there was an appeal to the fact that in human beings, there must be a neural signal that reaches the hippocampus in order for an episodic memory to be formed. This is, clearly, a fact (if it is

one) about types: events of the type 'neural signal reaching hippocampus' are causally necessary for events of the type 'forming a memory of an episode in one's personal history'. So how could it be appealed to in the present context?

The answer is that nothing like an identity has been presupposed. All that has been presupposed is that a certain generalization holds of enough human beings that we can be sure that it holds of Farhad in the situation imagined. This particular generalization does not violate the principles of any behaviourist, interpretationist, or a priori functionalist, because it is not held to be a necessary truth even about contemporary human beings, much less about episodic memory in any conceivable agent. The generalization is merely a generalization about biologically normal contemporary human beings who have lived ordinary lives free from radical neural insult. Behaviourism, interpretationism, and causal-role functionalism all *allow* it to be true that today neural signals reaching my hippocampus are required for my remembering some event, while tomorrow neural signals reaching my amygdala (and not my hippocampus) are required for remembering a qualitatively similar event. However, they cannot *require* that there is such a large degree of functional variability in actual people. And in fact, when it comes to episodic memory, this is never the way things work in actual people.[17] In people (indeed, in all mammals), neural signals reach the hippocampus or a new episodic memory is not formed.

How far do these facts about the hippocampus take us?[18] There are regions of the brain that show substantial variability. It appears that the fusiform gyrus of the cerebral cortex, for instance, is differently structured in people with different sorts of expertise: allowing car experts to recognize cars, bird experts to recognize birds, chess experts to recognize chess positions, and sighted people generally to recognize human faces (see Gauthier *et al.* 2000; Bilalić *et al.* 2011). For all we know, it could be that the same smallish anatomical structure in one person at one time in her life realized her ability to recognize a famous chess opening, and at a different time in her life realized her ability to recognize her grandchild's face (though of course that same anatomical structure would have to be differently functionally poised, by being differently strongly connected to other neural regions, in order to play the two different roles). Knowing that neurons in the fusiform gyrus played a role in an action would not amount to knowing much. Does this sort of variability threaten the importance of the causal map?

There are three responses to give. First, it is worth noting that the question already presupposes my main conclusion: that the philosopher of action or moral psychologist must not take for granted, but be informed of the

[17] At this point in history; perhaps, we will all get technological memory enhancements in the future, and the hippocampus will never be used again.

[18] Thanks to an anonymous referee for raising this question.

empirical facts to ensure, that she avoids inconsistency with the causal map. The question simply asks whether it might not be particularly easy to avoid such inconsistencies, given that some neural structures play different roles on different occasions, i.e., given some actual empirical facts. Secondly, the right stance in light of neural variability is to hold that any philosophical use of the causal map would have to be cautious, and take into account how much variability exists in how the brain creates, realizes, or grounds the mind in each region of the brain. Lessons learned from a single patient, or from American undergraduates of European descent, cannot be generalized to all human beings without further argument, for instance. But, thirdly, for all this caution, there is no reason to be sceptical about the frequency of generalizations such as the above generalizations about the hippocampus. In ordinary, healthy people, the hippocampus does one sort of job, the central nucleus of the amygdala does a different sort of job, secondary visual cortex does a third sort of job, the supplementary motor cortex does another job, the ventromedial prefrontal region of the cortex does yet another job, and so on.[19] Many investigations, of human beings and of other mammals, over many years, and across many experimental and disease conditions, show that in normal, adult human beings these structures and many, many others play stable functional roles throughout adulthood. Severe injury to area V1 of the adult visual cortex impairs almost all visual capacities permanently in human beings; severe injury to the ventral stream of visual processing (in the temporal lobe) creates different characteristic visual impairments, and severe injury to the dorsal stream of visual processing (in the parietal lobe) creates different characteristic visual impairments.[20] And so on. The current state of neuroscience strongly suggests that the causal map will impose a large number of causal constraints on philosophical theorizing, so long as philosophers have (as they appear to have) many causal commitments in their action theory and moral psychology.

VI. OBJECTION: THE FIRST-PERSON PERSPECTIVE

The final purely philosophical objection I will consider to the relevance of the causal map is inspired by Christine Korsgaard's work and the passage with which I began this paper. Korsgaard has long argued that what matters to moral psychology (and more) is the first-person perspective, not the third-person; if there is a conflict between the two, then what is visible only from the third-person perspective (a part of the Scientific World View, as she often calls it) is what is to be discarded. Recall her position:

[19] See Kandel *et al.* (2013) for canonical textbook treatments of these and many other stable generalizations about roles played by neural structures.

[20] Jacob and Jeannerod (2003) provide a philosopher-friendly discussion of this and many related phenomena, and at a finer grain than the present discussion.

> The freedom discovered in reflection is not a theoretical property which can also be seen by scientists considering the agent's deliberations third-personally and from outside. It is from within the deliberative perspective that we see our desires as providing suggestions which we may take or leave. You will say that this means that our freedom is not 'real' only if you have defined the 'real' as what can be identified by scientists looking at things third-personally and from outside. (Korsgaard 1996)

Setting aside the specific merits of this as a defence of freedom, it should be asked whether there is a strategy here that can be generalized to be used in the present discussion. Suppose that a philosopher were to claim that Farhad can know from the first-person perspective that he raised his hand because he thought it best, and not even partly because he had a standing desire (understood as a distinctive, non-cognitive attitude tied to pleasure, thirst phenomena, and so on) to do what he thought best. Or, if this is too much, then at least that Farhad can know from the first person perspective that he often does things because he thinks them best and not because of his standing desires. (This is meant to be a continuation of the previous example rather than a specifically Kantian variant, although of course a Kantian variant could be constructed.) And suppose that the causal map were to reveal an apparent internal inconsistency in commitments: suppose that the causal map were such that it was necessary to appeal to something having the causal powers (or something realizing or grounding something having the features) of a standing desire. Suppose this were true for every action of the sort the philosopher wanted to say was performed solely (or at least often) because of what was judged best. Would Korsgaard's thinking be of help to him?

Korsgaard's thinking seems most helpful in contexts in which there might be something missing from the science, by the scientists' own lights. If, for instance, the causal map were radically incomplete but appeared, at its current stage of development, to bar Farhad from acting solely out of a judgement that raising his hand is best, then the philosophical arguments that appear to make it plausible that Farhad can nonetheless act solely out of a normative judgement also seem at least moderately credible arguments that there is a fault in the present version of the causal map. To this extent, at least, a very weak form of the above Korsgaardian line of argument is reasonable. When we are sure something is real, and an incomplete science has not yet recognized necessary conditions for its existence, we are often justified in being at least fairly certain that the fault lies in the incomplete science, and not in our grasp of reality.

As the causal map reaches completion, though, this style of argument becomes much less credible. If a completed causal map were to imply that neural structures token-identical to (realizing, grounding, etc.) standing desires were always in fact causally involved in the events identified by everyone as Farhad's actions, how much room is really left for the theorist to object? It is the theorist's own conception of what standing desires must be like, to be distinctive non-cognitive attitudes, and the theorist's own conception of what conscious

judgements of what is best must be like, to be genuinely action-guiding, that have generated this problem. Perhaps the neuroscientists have missed a neural fibre tract or missed a mode of causal influence from one group of neurons to another, of course. But soon enough, when the domain is low-level neuroanatomy and neurophysiology, this sort of speculation is the equivalent of speculating that perhaps the thigh bone is not connected to the hip bone. And when push comes to shove, philosophical arguments about the nature of agency are much weaker than scientific arguments about the connection of the thigh bone to the hip bone.

The philosopher can, at this point, insist that what we take to be true from the first-person perspective *must* be taken to be true, or else the existence of personhood, agency, and the like cannot be sustained. He might take his inspiration here again from Korsgaard (2009), asking '[a]re the teleological and moral conceptions of the world then related to the Scientific World View as illusions to fact? If so, whose illusions would they be?' But what the philosopher is, in essence, claiming is that, if the science remains recalcitrant, he will declare that we must either embrace known empirical falsehoods or we must declare persons and agents to be non-existent. What seems much more likely, at this point, is that the philosopher has mischaracterized conscious thought, or mischaracterized standing desires, or made a mistake regarding the necessary conditions for personhood, agency, and the like. This is more likely because, as I should perhaps repeat, it is the philosopher himself who has generated the internal inconsistency. In this situation, it is not the criticism levelled by an external scientific world view, but the criticism levelled by internal commitments to things that cannot all be true, that is causing the problem for the philosopher.

VII. EMPIRICAL REASONS FOR PHILOSOPHERS TO IGNORE THE NEUROSCIENCE OF MOVEMENT

It seems, then, that even the philosophers who have held that their views are insulated from the causal map are nonetheless committed to it having or lacking certain features. Does it follow that these philosophers must now turn to the causal map for theoretical guidance?

The causal map does not yet exist. That is, an exhaustive anatomical and physiological account of the production of human bodily movement, explaining every (uncontroversially) measurable aspect, does not yet exist. It will not exist tomorrow either. This makes possible at least two reasonable objections to attempting to use the neuroscience of movement, that is, the current best indications regarding the causal map, in philosophy: one from the instability of incomplete science, and one from the incompleteness of current knowledge, even if it is granted that what is known is very likely stable.

First, one might object that current findings are not sufficiently stable. They might be the findings of our best current science, but there is a great deal left to learn, and future findings might overthrow present findings. Action theorists and moral psychologists are better off waiting until the causal map settles into a stable form.

The objection applies with justice to various research programs in the sciences. Even within normal science (Kuhn 1962) advancing normally, the objection can apply. For example, the role of cytochrome c in apoptosis (programmed cell death) was at the early stages of investigation in the mid-1990s (see Kroemer *et al.* 1998, to pick one example more or less at random). It would have been premature back then to leap to conclusions that, twenty years later, are now well established. Even within normal science, it generally takes years of work by multiple groups to reach stable insights. And then, there is the threat of a major theoretical revolution. For example, quantum mechanics famously appears inconsistent with general relativity, and so fundamental physics appears to be waiting for at least one more major revolution in a long line of major revolutions (see, e.g., Smolin 2001).

Regarding the instability of normal neuroscience, philosophers are of course well advised not to rely on single sources or the most recent research findings, unless these are very conservative extensions of what is already decade-plus old conventional wisdom about (low-level) neuroanatomy and neurophysiology. But the textbooks on these subjects are rich with information, and are not undergoing regular revolutions. That is, there is a large, so-far stable body of neuroscience relevant to the causal map that is available right now.

What is more interesting is the question of whether philosophers can set aside the findings of neuroscience because of the threat of major scientific revolutions to come. The question is not answered by noting that fundamental physics still awaits a revolution, since there are large branches of the natural sciences that, for good reasons, have not experienced revolutions since reaching their modern forms, and are very unlikely to experience revolutions in the future. General human anatomy and human physiology are two such disciplines. It took a revolution in human physiology to discard the four-humour theory, but that was quite long ago. Since the start of the twentieth century, although there have been many major discoveries in gross human anatomy and physiology, they have not had the revolutionary character of the overthrow of the four-humour theory. In spite of the necessarily tentative nature of all scientific findings, the knee-bone stubbornly remains connected to the thigh-bone in anatomy textbooks. Veins are still held to carry blood toward the heart, and arteries away from it. The stomach remains a key player in the theory of digestion. And so on.

The anatomy and the low-level physiology of the brain have, likewise, been normal sciences for between fifty years (neurophysiology) and a hundred

years (neuroanatomy). The foundational claims of these disciplines early in the twentieth century have not been overthrown by revolution. We still hold that the brain is composed of billions of distinct cells, not a continuous web (Finger 1994). We still hold that the great majority of influence had by these cells upon each other is mediated by passing chemicals from the axon of one to the dendrite of another, across the tiny gaps of the synapses (Valenstein 2002). Instead of revolution, these foundations have been added to, in a mainly cumulative pattern.

Perhaps, this non-revolutionary accumulation has happened because of the relatively straightforward techniques that have been used to advance the disciplines. The main techniques involve using dyes and looking through microscopes (in the case of neuroanatomy), and using electrical or chemical stimulation of neuron A while monitoring changes in neuron B (in the case of neurophysiology). There are fancier techniques as well, but the backbone techniques in these fields are as straightforward as one could ask.[21]

Higher level anatomical claims (e.g., that there is a meaningful unit that can be called the 'limbic brain') and higher level physiological claims (e.g., that the role of the limbic brain is to generate emotional responses to stimuli) are another matter altogether.[22] These claims cannot be substantiated by observation of what is under a microscope, or by detecting the release of GABA in location A after applying a tiny dose of glutamate to location B. They are, by their nature, more fraught and more likely to be overturned by later research. But they are also not findings that are involved in constructing the causal map. They are scientific interpretations of the current draft. So they do not concern the present discussion. The question is not whether philosophers should take seriously the higher level interpretations that neuroscientists have offered of the causal map, but whether they should take seriously the causal map itself.

Is this perhaps too sanguine? Philosophers following current cognitive science will know that there are strongly externalist (embodied) and dynamical theories of the mind currently prominent (early entrants here include Clark 1997 and Port and Van Gelder 1995). Do these theories not threaten revolution?

If successful, these theories threaten revolution at a higher level of abstraction than the causal map. Strongly externalist theories of mind take mental processes to extend well beyond the body, but (like contemporary Wittgensteinians) they do not deny that things inside the brain remain particularly important to the mind, and so do not deny the relevance of the causal map. At worst, they argue that there are vital but neglected extracranial elements

[21] It is important to not overstate how stable the neuroscience is, however. Saunders *et al.* (2015) reveals a brand-new neural pathway of influence from an important sub-cortical structure to an important cortical structure, for instance. Such discoveries are now fairly rare, but important to remember nevertheless.

[22] Morgane and Mokler (2006) provide one recent gateway to this debate.

that would be on any ideal causal map, meaning that consistency with the intracranial causal map would be a necessary achievement for an action theorist or moral psychologist, but not sufficient. And dynamical systems theorists of the mind hold that certain kinds of causal systems (implementing dynamical feedback loops) are of particular importance to the mind, and require special new ways of understanding them at higher levels of abstraction. At worst, they argue that the important parts of a map of cause and effect will have many complex feedback loops within them. But that would not undermine the claim that a correct causal map, showing these feedback loops, would be something that the action theorist or moral psychologist would need to take into account, and might fail to take into account properly. That there are many feedback loops in a system does nothing to show that all causal claims are true of that system!

Finally, one might reasonably worry about recent work done by Michael Anderson attacking modularity. Anderson (2014) synthesizes and adds to the evidence that higher level neurophysiology still requires a revolution, one that will produce an understanding of the brain as much less modular than currently accepted. And Anderson is certainly not alone. But even if such a revolution is carried out, it would not require any changes to the low-level facts that are displayed in the causal map. It would, however, suggest that interpreting the causal map will be harder even than currently believed. Note, though, that such a revolution would not touch the in-principle points made earlier that there is no way for a philosopher to escape causal commitments. Even if those causal commitments are commitments to features of a minimally modular, highly dynamic, mathematically complex (in the sense of complexity theory) system, they are still commitments to different, and possibly conflicting, such features. Note also that Anderson himself is sceptical that there will be a revolution regarding the immediate production of action; he sees his revolution as taking place largely in the cognitive sphere.[23]

Thus, while revolution is always a possibility in any branch of science, it is not a possibility any philosopher should take seriously in the gross anatomy or physiology of the human body as a whole. The anatomy and physiology of the brain, at the level of focus relevant to producing the causal map, are not quite as secure, but not by much: they are similarly resilient, and so provide similarly weak reasons to ignore what is now known. The various revolutions that might yet arrive in the study of the mind are all revolutions at higher levels of analysis than that of the causal map.

A second objection that might be raised about the incomplete causal map is that, because it remains incomplete, it would be premature for a philosopher to attend to the current draft.

[23] Anderson, pers. comm.

This objection relies on ignorance of the actual details of current drafts of the causal map. At the level of grain that is clearly relevant to the questions mentioned earlier—questions such as whether canonically voluntary bodily movements have three different sorts of causes (corresponding to appetite, passion, and reason) or one (for instance, reason)—the causal map is as close to complete as it needs to be and has been so since (at least) the start of the twenty-first century. A range of textbooks, from general graduate texts in neuroscience (e.g., Kandel *et al.* 2013) to anatomical atlases (e.g., Hendelman 2000; Rosenberg 1998) to texts focused on movement (e.g., Riehle and Vaadia 2004; Rothwell 1993) and movement disorders (e.g., Fahn, Jankovic, and Hallett 2011; LeDoux 2014) provide textbook-grade explanations of the causal map at the level of resolution that action theorists and moral psychologists are likely to care about. No neuroscientist has yet produced exactly the textbook that philosophers would most like to have: a textbook that would describe the causal map with a focus on the interests of philosophers, and in language suited to philosophers. So pointing to these textbooks is not meant to suggest that philosophers have been lazy or negligent. There is hard work ahead for a philosopher interested in interpreting the causal map. But the point is that the science, at the level of detail relevant for doing philosophy, is largely completed.

Just what does this science say? Answering this question is the task of a book, not a paragraph. But in outline, it appears to me that movements of the sort most commonly regarded as paradigmatic voluntary movements in human beings are the immediate product of three different kinds of neural influences. Influences from neurons that are primed to immediately cause simple bodily movements (one might say, influences from the basic actions one is ready to perform) join with influences from neurons from a very wide array of apparently sensory and apparently cognitive regions of the brain (one might say, influences from how things seem to be) and influences from a small group of neurons, releasing dopamine, and tied to causing immediate impacts on movement, feelings, and patterns of attention (one might say, influences from what one wants). These three influences meet deep in the brain, and produce the release of certain of the already-primed bodily movements (while continuing a general restraint of all other possible bodily movements): choices, one might say, of the basic actions one will perform at each moment.[24]

If this description raises more questions than it answers, it has had its hoped-for effect. The causal map is inescapably relevant to questions philosophers of action and moral psychologists care about, and complete enough for us to consider. It is now time to understand for ourselves what it means.[25]

[24] See Schroeder (2004). Compare to Yaffe (2013), Holton (2009), and Railton (2014).

[25] This paper benefited from the comments I received on early versions of it presented at Rice University, the University of Arizona, and the University of California at Riverside, and to the members of the Moral Psychology Research Group. I also received very helpful comments from Nomy Arpaly and two anonymous referees.

REFERENCES

Anderson, M. (2014) *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA: MIT Press.

Arpaly, N. and Schroeder, T. (2014) *In Praise of Desire*. New York: OUP.

Bennett, M. and Hacker, P. (2003) *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.

Berker, S. (2009) 'The Normative Insignificance of Neuroscience', *Philosophy and Public Affairs*, 37: 293–329.

Bilalić, M. et al. (2011) 'Many Faces of Expertise: Fusiform Face Area in Chess Experts and Novices', *Journal of Neuroscience*, 31: 10206–14.

Brandom, R. (1994) *Making it Explicit: Reasoning, Representing, and Discursive Commitment* Cambridge, MA: Harvard University Press.

Butler, K. (1992) 'The Physiology of Desire', *The Journal of Mind and Behavior*, 13: 69–88.

Churchland, P.S. (1986) *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: MIT Press.

—— (2002) *Brain-wise: Studies in Neurophilosophy*. Cambridge, MA: MIT Press.

Churchland, P. M. (1995) *The Engine of Reason, the Seat of the Soul*. Cambridge, MA: MIT Press.

—— (2007) *Neurophilosophy at Work*. New York: CUP.

Clark, A. (1997) *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Davidson, D. (1980) *Essays on Actions and Events*. Oxford: OUP.

—— (1984) *Inquiries into Truth and Interpretation*. Oxford: Clarendon.

Davis, W. (1986) 'The Two Senses of Desire', in J. Marks (ed.) *The Ways of Desire: New essays in Philosophical Psychology on the Concept of Wanting*, 63–82. Chicago: Precedent.

Dennett, D. (1987) *The Intentional Stance*. Cambridge, MA: MIT Press.

Dill, B. and Holton, R. (2014) 'The Addict In Us All', *Frontiers in Psychiatry*, 5: 1–20.

Fahn, S., Jankovic, J. and Hallett, M. (2011) *Principles and Practice of Movement Disorders*, 2nd edn. New York: Saunders.

Finger, S. (1994) *Origins of Neuroscience: A History of Explorations into Brain Function*. New York: OUP.

Gauthier, I. et al. (2000) 'Expertise for Cars and Birds Recruits Brain Areas Involved in Face Recognition', *Nature Neuroscience*, 3: 191–7.

Greene, J. (2010) 'The Secret Joke of Kant's Soul', in T. Nadelhoffer, E. Nahmias and S. Nichols (eds) *Moral Psychology: Historical and Contemporary Readings*, 359–72. Oxford: Wiley-Blackwell.

Hendelman, W. (2000) *Atlas of Functional Neuroanatomy*. Boca Raton: CRC Press.

Holton, R. (2009) *Willing, Wanting, Waiting*. New York: OUP.

Hornsby, J. (1997) *Simple Mindedness: In Defence of Naive Naturalism in the Philosophy of Mind*, Cambridge, MA: Harvard University Press.

Jackson, F. and Pettit, P. (2004) 'Moral Functionalism and Moral Motivation', in F. Jackson, P. Pettit and M. Smith (eds). *Mind, Morality, and Explanation: Selected Collaborations*, 189–210. New York: OUP.

Jacob, P. and Jeannerod, M. (2003) *Ways of Seeing: The Scope and Limits of Visual Cognition*. Oxford: OUP.

Kandel, E. et al. eds (2013) *Principles of Neural Science*, 5th edn. New York: McGraw-Hill.

Korsgaard, C. (1996) *The Sources of Normativity*. New York: Cambridge University Press.

—— (2009) *Self-Constitution: Agency, Identity, and Integrit*. New York: Oxford University Press.

Kroemer, G., Dallaporta, B. and Resche-Rigon, M. (1998) 'The Mitochondrial Death/Life Regulator in Apoptosis and Necrosis', *Annual Review of Physiology*, 60: 619–42.

Kuhn, T. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

LeDoux, J. (1996) *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.

LeDoux, M., ed. (2014) *Movement Disorders: Genetics and Models*, 2nd edn. Amsterdam: Academic Press.

Levy, N. (2014) 'Addiction as a Disorder of Belief', *Biology and Philosophy*, 29: 337–55.

Mölder, B. (2010) *Mind Ascribed: An Elaboration and Defence of Interpretivism*. Philadelphia: John Benjamins.

Morgane, P. and Mokler, D. (2006) 'The Limbic Brain: Continuing Resolution', *Neuroscience and Biobehavioral Reviews*, 30: 119–25.
Morillo, C. (1990) 'The Reward Event and Motivation', *Journal of Philosophy*, 87: 169–86.
Port, R. and van Gelder, T. (1995) *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.
Railton, P. (2014) 'The Affective Dog and Its Rational Tale: Intuition and Attunement', *Ethics*, 124: 813–59.
Riehle, A. and Vaadia, E., eds (2004) *Motor Cortex in Voluntary Movements: A Distributed System for Distributed Functions*. Boca Raton: CRC Press.
Rosenberg, R., ed. (1998) *Atlas of Clinical Neurology*. Newton, MA: Butterworth-Heinemann.
Roskies, A. (2003) 'Are Ethical Judgments Intrinsically Motivational? Lessons from "Acquired Sociopathy"', *Philosophical Psychology*, 16: 51–66.
Rothwell, J. (1993) *Control of Human Voluntary Movement*, 2nd edn. London: Chapman and Hall.
Ryle, G. (1949) *The Concept of Mind*. London: Hutchinson.
Saunders, A. et al. (2015) 'A Direct GABAergic Output from the Basal Ganglia to Frontal Cortex', *Nature*, 521: 85–9.
Schroeder, T. (2004) *Three Faces of Desire*. New York: OUP.
Schueler, F. (1995) *Desire: Its Role in Practical Reason and the Explanation of Action*. Cambridge, MA: MIT Press.
Schultz, W. (2015) 'Neuronal Reward and Decision Signals: From theories to data', *Physiological Reviews*, 95: 853–951.
Shea, N. (2014) 'Reward Prediction Error Signals are Meta-Representational', *Noûs*, 48: 314–41.
Smolin, L. (2001) *Three Roads to Quantum Gravity: A New Understanding of Space, Time, and the Universe*. New York: Basic Books.
Stich, S. (1996) *Deconstructing the Mind*. New York: OUP.
Valenstein, E. (2002) 'The Discovery of Chemical Neurotransmitters'. *Brain and Cognition*, 49: 73–95.
Walter, H. (2001) *Neurophilosophy of Free Will: From Libertarian Illusions to a Concept of Natural Autonomy*. C. Klohr (trans.) Cambridge, MA: MIT Press.
Wittgenstein, L. (1953) *Philosophical Investigations*. E. Anscombe (trans.). Oxford: Blackwell.
Yaffe, G. (2013) 'Are Addicts Akratic? Interpreting the Neuroscience of Reward', in N. Levy (ed.) *Addiction and Self-Control: Perspectives From Psychology, Philosophy, and Neuroscience*, 190–213. New York: OUP.

Rice University, USA