Criminal Law and Philosophy (2022) 16:327–351
https://doi.org/10.1007/s11572-021-09600-w

ORIGINAL PAPER

# Neuroscience and Normativity: How Knowledge of the Brain Offers a Deeper Understanding of Moral and Legal Responsibility

**William Hirstein**[1]

Accepted: 16 July 2021 / Published online: 27 July 2021


## Abstract

Neuroscience can relate to ethics and normative issues via the brain's cognitive control network. This network accomplishes several executive processes, such as planning, task-switching, monitoring, and inhibiting. These processes allow us to increase the accuracy of our perceptions and our memory recall. They also allow us to plan much farther into the future, and with much more detail than any of our fellow mammals. These abilities also make us fitting subjects for responsibility claims. Their activity, or lack thereof, is at the heart of culpability. For instance, planning to kill someone is strong evidence of what the law calls men rea—a guilty mind. Claims about norms, or ethical "should" claims, express two-level propositions, directed at the behaving person at one level, and at that person's mind and cognitive control network at another level. Thus, "People should stop themselves from hurting others," is a claim about how people should behave and also a claim about how their cognitive control networks should behave—i.e., they should inhibit harmful behavior, or the intentions leading up to it. Planning is both an ability of the full person, and of that person's mind. Neuroscience affirms the common notion, seen both in law and folk psychology, that what makes us guilty or culpable are certain events and states that exist in our minds. Overt behavior, including speech, is fallible evidence of these states and processes. Cases of negligence still involve the executive processes, but "negatively," in that negligence results when certain types of executive activity fail to take place.

**Keywords** Responsibility · Executive processes · Cognitive control · Culpability · Inhibition

✉ William Hirstein
williamh@elmhurst.edu

[1] Elmhurst University, Elmhurst, USA

## 1 Introduction

Before we hold people responsible for wrongdoings, we want to know what went on in their minds leading up to the act. We want to know what their plans and intentions were, and how they were executed. This knowledge can strongly affect our judgments about how culpable that person is. Contemporary neuroscience is accumulating a vast store of findings about the brain, but is there anything among them that can tell us more about the mental features of responsibility and culpability? In our book, *Responsible Brains: Neuroscience, Law, and Human Culpability* (RB) (2018), Katrina Sifferd, Tyler Fagan, and I constructed an account of responsibility based on neuroscientific findings. We argued there that the hypotheses and experimental findings of neuroscience are a valuable guide in forming theories of responsibility. Our primary point was that the brain's cognitive control network and its executive processes—which include attention, task switching, the monitoring of perceptions, memories and emotions, the monitoring of behavior, manipulating items in working memory, planning, and inhibition—are quite relevant to our understanding of the mental components of responsibility. We welcome this chance to respond to feedback from several eminent experts on these problems.

This paper contains replies to the six commentators in this issue: Craig Agule, Federica Coppola, Douglas Husak, Michael Moore, Stephen Morse, and Dennis Patterson. Their responses range from corrections to the details of our account, to sweeping indictments of our entire project. I will begin with the latter, by addressing doubts the commentators expressed about the normative or ethical import of neuroscience. Then I will address several objections raised to our treatment of specific problem cases in the theory of responsibility, including cases involving schizophrenic defendants, "sleepwalking" cases, and negligence cases.

Four of the six commentators express serious doubts in their responses about the relevance of neuroscience to our understanding of responsibility. According to Husak, neuroscience has no "consequences for our judgments of responsibility," and fails to "*inform* the assessments of culpability we would otherwise make" (2021, this issue). Patterson worries that we, the authors, "believe that the brain can inform our judgments of responsibility for action." "Neuroscience at present," according to Morse, "contributes almost nothing to the necessary psychological level of explanation and analysis." Coppola is also unmoved: "the Authors' theory ultimately adds little to the dominant understanding of responsibility in legal theory and doctrine" (2021, this issue).

The claim of irrelevance is made specific in four different ways by the commentators, each of which is treated in a subsequent section:

1. The information we have from neuroscience consists only of the names of brain areas, the activity of which was found to correlate with certain behaviors or situations. Correlation is not causation, hence this information is disconnected from the causal nexus we are interested in—actions for which people can be held responsible (Morse).

2. Even if the neuroscience is correct, it adds nothing to our ongoing study of responsibility. Mere knowledge about the brain events behind our actions can tell us nothing about whether people were or were not responsible for those actions (Husak).
3. In general, neuroscience has no normative consequences, and so cannot shed light on a concept riven with normativity such as responsibility (Patterson, Husak).
4. Neuroscience is parasitic on psychology and knowledge of behavior for any significant connections it has to responsibility (Morse).

## 2 Objection: Neuroscience Provides Only Names and Correlations

According to this objection, findings from neuroscience are merely "correlational," in the sense that we see brain areas become active via fMRI when someone is behaving a certain way, or perceiving a certain stimulus, but we don't understand the causal links between the brain activity and the behavior. We could always be mistaking mere correlation for causation. "As a result," says Morse, "all language indicating causation, such as a region enabling or supporting task behavior, is loose talk" (this issue).

This problem is exacerbated, if not caused outright, by reports of brain findings in the philosophical literature that name an area found to be active, while failing to contextualize that information by describing the area's anatomy and physiology, as well as other relevant information. If all that is being provided is the name of a brain area, with no indication of its functions, its connections to perceptual areas, its connections to areas involved in generating behaviors, its connections to other brain networks, its neurochemistry, the consequences of damage to that area, how that area develops in childhood and declines in senescence, what structure and activity that area shows on the myriad types of brain imaging available to us, and so on, the skepticism is understandable.

The claim that our ability to connect the brain to behavior is merely correlational is no longer true, however. For the most part, we understand the brain's causal nexus. Through cell-staining studies and more recently through the use of diffusion tensor imaging, we largely understand the nature and extent of the brain's connections. For instance, we can observe activity in one cortical area, followed by activity in its white matter connecting fibers, followed by activity in another area. Granted, there is still a possible gap here, in that there could be another causal connection we do not know about that is causing the subsequent activity. But the possibility that a significant connection could be hiding from our many techniques grows vanishingly small as research progresses.

Our knowledge about the brain is coalescing from several different subdisciplines as it also incorporates and re-interprets information from classical neurology about brain lesions and their consequences. We now have our first large-scale theories of how the cortex operates at the top level. It is divided into several networks, including the cognitive control network, consisting of interconnected areas and their supporting subcortical areas (Yeo et al. 2011; Eickhoff et al. 2018). There are also now

many other brain imaging techniques in addition to MEG, fMRI, and PET. This multiplicity of approaches is crucial to the success of neuroscience since it allows cross-checking; fMRI is the most popular, but it relies on cross-checking against other brain imaging and recording techniques.

All of this offers us a vast new framework—neuroscience and its related disciplines—with which to pose new questions and make new discoveries about the strengths and weaknesses of our reasoning apparatus, the way it responds to different types of damage, the nature of its links to the brain's emotion systems, as well as its links to consciousness, all of which are relevant to understanding responsibility.

## 3 Objection: Neuroscience Adds Nothing to Our Understanding of Responsibility

Suppose we were to be granted complete knowledge of the mind of a defendant in a case in which their motives and intentions were unclear. This would surely be of great relevance and import to their guilt or innocence, especially in cases where *mens rea*[1] must be established. Knowing about the mind is in fact knowing about the brain, we would argue, and it seems most of the commentators agree. So then, if knowing about mind is of great import to questions of responsibility, and knowing about the mind is knowing about the brain, it follows, in some reasonably strong way, that knowing about the brain is of great import. Where is the slippage?

We are in the early stages of understanding how the executive processes operate, and how the cognitive control network achieves them. It is quite possible that specific information about how they operate could affect our conceptions of responsibility. When, for example, can a person inhibit an impulse and when can't they? What are the mechanisms by which inhibition occurs? Understanding the executive functions can help with the issue of how "resistable" emotions,[2] impulses, and obsessions are. Is powerful anger as difficult to resist as powerful fear? In other types of cases, we might find that we should hold each other less responsible. What we have learned about people with Tourette's syndrome, for example, tells us that it involves significant differences in the network activity in their brains, which may indicate that their ability to inhibit actions is compromised (see, e.g., Fan et al. 2018). This could mean that people with the condition should not be held responsible for certain sorts of apparent actions. In the future, we will be able to distinguish between actions that were produced abnormally, and those that were produced using the normal machinery of intentional action (see, e.g., Hampson et al. 2009). We might

[1] "*Mens rea*" ("guilty mind") refers to mental states or events, such as plans or intentions, that make a person culpable, given certain actions.

[2] This could help with the sorts of duress cases that Coppola mentions. Coppola is right that we neglect the emotions in *RB*, short of making the point that the executive processes are tasked with emotional regulation, for instance, taking care to be sure that we express the right amount of emotion given the situation. Certainly "emotion" and "cognition" interact a great deal. The brain processes that subserve them are heavily interconnected, but nevertheless anatomically separate, which allows us to diagnose problems as either emotional or cognitive, which could have legal implications.

find, for instance, that the actions in question are produced without the proper causal interactions between the areas responsible for self, consciousness, or rationality.

Learning more about the brain will inevitably force changes in law and in our conception of responsibility. For example, there is a set of laws concerning the age of adulthood, voting, consent, etc. Neuroscience can offer findings about when the areas required to have the capacities needed to make and be held responsible for such decisions mature. The set of scientific theories covering this development can eventually describe different developmental trajectories for the different abilities. "Does this information," asks Patterson, "tell us anything we did not already know from observation of the behavior of adolescents?" (this issue). It gives us much more specific information about what their executive weaknesses are. It informs us about the root causes of their moral immaturities, which might not be what behavior indicates they are. For example, their affective decision making is immature compared to their cognitive decision making, so that they perform similarly to Damasio's patient on tests of risky decision making (Crone and van der Molen 2004).

Our existing concept of responsibility contains specific requirements concerning how the actions that we are responsible for are generated. Neil Levy (2014) has made the point that our ability to be consciously aware of the situation and what is at stake is important to responsibility. The deep-self theories of responsibility correctly point out that actions we are responsible for have a sort of approval from us, in that they are consistent with our values (Doris 2017) and our desires, and even express them (Sripada 2016). People also need to possess the ability to be reasons-responsive in order to be held responsible (Fischer and Ravizza 1998). Each of these mental abilities, along with the ability to be responsible itself, is produced by a set of brain areas and networks, which of course exist in the larger context of a human body situated in a society.

According to both the folk and legal conceptions of responsibility, these three abilities—consciousness, sense of self, and rationality—need to be related in certain ways in order for human responsibility to exist. In *RB*, we describe which systems and processes are behind the above abilities, and how they relate. The executive processes can unify our understanding of these abilities since they connect all three. They produce a sense of self, and we use "I" when referring to them, and they function as a sort of stand-in for the person at the mental level. They also manage the process of assessing whether a contemplated action achieves our desires and expresses our values and beliefs, a process which makes use of the global conscious workspace. The executive processes also play a large role in making us reasons-responsive. Hence the executive processes play multiple vital roles in allowing us to be responsible beings who are fitting subjects of culpability. Consciousness also knits together all three of the other abilities (identity/self, rationality, responsibility). The arena of culpability, the locus of *mens rea*, is an executive-controlled conscious workspace. Since the workspace is global (Baars 1997), the resources needed to relate contemplated actions to one's values and desires, and to one's sense of self and identity as well as one's ability to reason, are also present on the conference call.

There are numerous other ways that these underlying processes need to interact in order to be aligned with our folk concept of responsibility. For instance, the right causal access must exist between the executive processes and memories, including

their value tags. Being aware of the values of people, things, and outcomes is crucial to making good decisions. The sort of control the executive processes have over the mental realm and over behavior should roughly align with our sense of which mental processes and actions are under our control and which are not. According to our understanding of current neuroscience these mental abilities do in fact relate in ways that are roughly consistent with both folk psychology and with a rough consensus found in our legal practices and doctrines—or so we argued in *RB*. Such a consistency could offer powerful affirmation of the folk and legal conceptions of responsibility.

### 3.1 Consciousness and Sense of Self

If a brain area causes significant (e.g., potentially harmful), intelligent actions without interacting with consciousness the person is less responsible for those actions. In *RB*, we argued that this is because the action did not receive "executive approval," which requires coordinated activity between the current conscious state and the executive processes, as well as other participants in the workspace system.

In our study of the brain, it is possible that we could find that what appeared to be intentional actions were generated outside of consciousness without any executive contribution. This is what Libet-type cases are alleged to show—that what we think of as intentional, voluntary actions happen automatically, even before we consciously intend to undertake them (Libet 1985). We argue in *RB* that Libet fails to notice that there is a conscious, intentional decision, involving executive activity, made prior to the first experimental trial, upon hearing the instructions. The subject decides to follow the experimental plan. The subject may then follow the plan in a somewhat automatic way, as we do with any action, but this does not imply that there was not an overarching, conscious decision. Another reason why Libet must be wrong is that we quite successfully treat certain behaviors as being under the control of people, and undertaken with their awareness and conscious consideration. The fundamental success of both folk psychology and the legal system, which are interconnected by thousands of conceptual threads, is based on something real: the brain and its activities. We simply need to interpret and describe the neuroscientific findings in such a way that their alignment with folk psychology and the law is clear.

Since consciousness is important to our issues here, we need to employ the best scientific theory of it. At the moment, however, there are two strong competitors, one of which emphasizes the primacy of phenomenal consciousness while the other uses access consciousness as its model (see Block 1995 on the distinction). In *RB*, we argued that what is commonly referred to as phenomenal consciousness is the best candidate for the process known as consciousness, for several reasons. The main reason is that the concept of access consciousness mistakenly includes the executive processes—the "access" occurs primarily between the (phenomenal) conscious state and the executive processes. This type of theory is thus describing a more sophisticated and complex type of brain state, involving, as they note, large cortical expanses in both the front and back of the brain (hence the "fronto-parietal" theory of consciousness). The other problem with this approach is that it makes the

traditional mind–body problem or the problem of consciousness insoluble. Binding persons so tightly to their conscious states creates an epistemic-metaphysical simple—an event that by its very nature can be known about directly by only one person—and is hence permanently "simple," rather than complex, in that it cannot be further broken up into two parts: consciousness and self—the subject and the conscious state that is its object. And there the story must end for those who espouse the access-based model of consciousness, without a satisfying analysis. Binding consciousness too closely to self/subject creates a realm of events that only the subject can know about, forcing a distinction that results in a dualist ontology (Hirstein 2012). They are defining consciousness as what I would call self-consciousness, a more complex state. No one else can experience your conscious state because you are inextricably bound to it, or so the dogma goes.

Michael Moore, like Levy, favors the access view, saying that he "wants to use consciousness to draw the line between processes that are 'personal' (things a person does) and processes that although they may necessarily underlie what a person does, are not themselves things a person does, what Dennett called 'subpersonal'" (2021, this issue). Moore continues, connecting the personal–subpersonal distinction with the distinction between phenomenal and access consciousness: "This is to use consciousness to draw the boundaries of a self. This is also to use 'consciousness' in its dispositional sense, not its phenomenal sense" (this issue). Moore mentions Dennett's view, which is similar to Levy's in that it binds consciousness to the self. But there are conscious states without a self, as in deep meditation, or focused concentration. There can also be conscious states with a self that is significantly different from our normal self, as in dreaming and dissociative identity disorder. The access-based view also implies false claims about the unattended portions of conscious states, such as the periphery of the visual field. No one, including the subject, is paying attention to them, yet they are still portions of visual conscious states.

The two opposing theories of consciousness figure heavily in our debate in *RB* with Levy's consciousness-based account of responsibility. Moore suggests that Levy might modify his position by accepting that executive function is necessary to responsibility, but arguing further that "so is dispositional awareness; executive control functioning where there is no privileged access to the processes of that functioning—such as the mechanisms of balance adjustment of the upright human body—are not things persons do or for which they can be held responsible" (this issue). It seems right that some access between the executive network and the relevant content is vital. But as I noted above, this conscious access is not necessarily privileged. Indeed, it is not clear that the specific process of consciousness is necessary to achieve the right sort of access (which I assume can only be causal access). This position may be right about *human* responsibility, but not for responsibility in general; there could be intelligent, rational alien beings, or autonomous robots that we would want to hold responsible for their acts, but whose "brains" did not contain conscious states of any sort. This would tend to show that the specific process of consciousness is not necessary, but that some broader category of causal access to

certain information is necessary, such as information about the situation, the being's own "values" and "desires" and identity, and so on. Our conscious global workspace[3] allows for the causal interactions, but they could be achieved in other ways. This might sound like a minor difference, but it shows that the executive processes are vastly more important to responsibility than (phenomenal) consciousness.

### 3.2 Rationality

Does the executive account have anything to offer by way of explaining our reasons-responsiveness or the sort of practical rationality that Moore points out is important to the issue of responsibility? The principal role of the executive processes in practical rationality is to ensure that the actions chosen are the best ones. They work at every part of the process of action production to ratchet up the success probability of the resulting action. They correct misperceptions, resolve contradictions in beliefs, and manage the process of assessing the values of potential action outcomes. They then stay online during action execution if needed, to quickly revise plans on the fly; i.e., they monitor.

At the heart of rationality lies logic. People who repeatedly violate basic logical principles cannot be said to be rational. Logic itself is often said to be normative, so it looks like there may be another gap here between neuroscience and the responsibility debates in philosophy. Logic is investigated a priori, while scientists work a posteriori. Another way to characterize this gulf is to say that our rationality partly involves our ability to do things for good reasons and "respond" to reasons. But the relationship between a reason for doing something and the conclusion that one should do that thing is logical, not causal. And brain science only deals in causal relations.

It is important that decisions have causal power. If the important features of decisions are logical, in that they involve the relations between reasons and conclusions, this might be taken to imply that they are not also causal emanations of physical things. How does one get from reasons to causes or back the other way? Logical relations can be captured by a system of causal relations, to a near enough degree of accuracy. Our computers do this. The computer on your desk uses silicon circuits to capture logical relations. It is reliable and accurate enough for us to productively assume that it is performing logic. But logic is perfect and no machine is perfect. The computer will one day fail to function. An effect will fail to follow its cause. One day there will be a one where, according to logic, there should (notice the normativity) have been a zero. Our computers are not perfect, but they are reliable enough. We ourselves mimic logical relations using causal relations. The brain frequently performs operations that could be described as implication—if–then. Our reliability level in such cases is lower than that of our computers—partly because

[3] To be clear, I accept Baars' idea that there is a global workspace that allows conscious contents to be widely "broadcast" to several modules. But Baars makes the same mistake that Levy and do by including the executive processes in his account of the neural locus of consciousness, calling it the "fronto-parietal" theory. See pages 103–108 of *RB*.

we are seduced by fallacies—but in rational people it is good enough for everyday purposes.

It is also important to note that there are scientific discoveries we could make that could threaten our existing notions of how our reasoning operates, and how it interacts with the brain processes that produce consciousness, identity, and ultimately responsibility. What if our reason-responsiveness or our ability to be rational was found to emanate from a small brain nucleus, whose workings were entirely inaccessible to conscious cognition? This would contradict the folk and the standard legal approaches. Both the decision process and its outcome need to be open to conscious view for us. We need to use the conscious workspace to assess values of possible choices and their outcomes. We also need to make sure that we get the outcome we desire. The decision process needs to reflect our beliefs, values, and desires at certain points. The conscious global workspace allows us to achieve this, partly by allowing thoughts held in consciousness to generate value tags from the brain's emotion and reward centers.

Future progress in neuroscience could reveal that the neural underpinnings of consciousness, our sense of self, and our rationality do not interact in the ways we imagined, or worse, that they do not interact at all. This could ultimately have profound effects on the ways we conceptualize responsibility, and on our practices of holding one another responsible.

### 3.3 Responsibility and Water

Another sort of response to Husak's claim that neuroscience adds nothing to the discussion is that a discovery could confirm or corroborate our folk understanding of a phenomenon in an important way without correcting that understanding or even adding to it at the folk level. The discovery that water is $H_2O$ did not contradict or eliminate any of our existing folk beliefs about water. We always believed it was a unique, homogeneous substance, capable of mixing with other fluids, in rivers, lakes, and oceans, vital to life, etc. What does knowing that water is $H_2O$ add to our current scientific understanding of water? It connects water to another discipline. It tells us what water is made of, and an endless list of things follows from this: That water contains hydrogen. That it contains oxygen. That it will relate in certain ways to certain other chemicals. That it can perform certain functions in biological systems. It was not apparent to anyone that water is $H_2O$ for the vast majority of human history, and without the field of chemistry, that finding would never have been made. In *RB*, we are saying that the human mental requirements for being responsible and culpable are "made of" the executive processes, interacting with other brain areas. This was not apparent to anyone prior to the discovery of executive processes.

Neuroscience provides a completely different framework, based in the sober and thoroughly objective science of biology, to provide a check against folk psychology and legal thought and practices, as well as our philosophical accounts of responsibility. In *RB*, we are reporting that we have assembled the relevant parts of neuroscience into such a framework, held it against folk psychology and the law, and found that there is fundamental agreement, despite several interesting discrepancies. In

order to provide this affirmation, the neuroscience had to be collated and interpreted correctly. It can then be used to assist our theory building in philosophy. The mistakes made by Levy and Libet—if the analyses above and in *RB* are correct—about the nature of consciousness and its relation to our intentions, respectively, caused them to give faulty analyses and explanations of how intentional actions work, in Libet's case, and in Levy's case, of the culpability of people such as Kenneth Parks, who commit crimes under the influence of REM behavior disorder (see the analysis below). If neuroscience can be used to correct competing theories of intentional action or responsibility, then it is relevant to those issues.

## 4 Objection: Neuroscience has no Normative Significance

Dennis Patterson says that "facts about the brain have no obvious normative implications" (2021, this issue). He quotes Berker approvingly: "Either attempts to derive normative implications from these neuroscientific results rely on a shoddy inference, or they appeal to substantive normative intuitions … that render the neuroscientific results irrelevant to the overall argument" (Berker 2009, p. 294). Husak notes that our criteria for responsibility in RB contain a use of that important term, "should" in criterion 3: Jo's executive processes either played the appropriate role in bringing about a consequence, or *should have* played an appropriate role in preventing it. This might be seen as sneaking an unexplained normative element into our account. The account not only contains a "should," but applies that "should" to brain processes. Husak is right—we do have more explaining to do about the role of normativity in the executive account, while avoiding the two errors mentioned by Berker. Below, I will provide a rough sketch of how we might do that.

### 4.1 Social Functions

Societies can be fruitfully thought of as functional systems. Herbert Spencer (1860) suggested that a society is like a living biological organism. The parts of society are like the organs of the body in that they engage in a complex set of functional relations that forms a unified system. The bodily organs, by analogy, Spencer saw as social institutions, such as religion or the educational system. Individual persons are functional units at the next level down, one would think, perhaps analogous to cells. Just as with every part of a functional system, there are things that people should and should not do, if the society is to operate well and flourish. These social functions need not be situated within pre-formed goal structures. This could be seen as importing an undefended normative element, in this case a teleological one. There are historical accounts available that do not require forward-looking components. For example, social functions are those features of humans that played certain (functional) roles, which were selected for (Wright 1973; Milliken 1989; Neander 1991), and then further shaped by selection according to how well they performed these selected functions. The social functions can be spoken of as if they had a goal, for

convenience, but the evolutionary process itself has no forward-looking component (unless one counts us humans and our goals).

In actual societies, the social functions are partly implemented by an enforced set of social norms. Social functions can be stated more generally than social norms. Societies share the social functions, while norms vary among societies. For example, one social function of humans is to form families. Most societies use a social norm of getting married in a certain sort of ceremony to support this social function, but there is great variety. Social functions and social norms overlap heavily, at least in healthy societies. The functions promote the health of the society, whereas norms need not. There are repressive social norms, for example, such as the caste system in India, that are harmful to vast portions of a society. Apparently, there are essential social functions, then we "derive" norms from them, sometimes incorrectly.

Whether or not a thing, process, or event performs a certain function is a factual matter. It is a fact that the heart performs the function of pumping blood. Thus, statements about functions, such as, "The heart functions to pump blood," are factual statements. Statements about functional failures are also factual statements. X *fails* to perform function y when and only when X *should have* performed y. A "should have" is just a "should" that the subject did not make happen. That is, to say that X should have performed y is the same as to say that X should perform y and X did not perform y.

Now we can complete criterion 3 and show the source of its normativity. Its full form is: "According to our [the human species'] social functions, Jo should have performed y (the appropriate role in preventing an event, etc.)," or "According to our [this society's] social norms, Jo should have performed y." Maintaining a distinction between social functions and social norms allows the account to avoid a relativism of responsibility that would result from tying it to social norms alone. For example, the account is able to say that a certain culture is mistaken in not holding people responsible for a given action and/or result (perhaps because no norm was violated). The culture can be mistaken if a social function is violated, because such functions are biologically ascertained features of the human social animal and not culturally relative.

### 4.2 Duplex Concrete Propositions

Consider the following list of truisms that describe commonly held social norms:

People should **plan** events, especially when they involve other people.

People should stop or **inhibit** themselves from performing contemplated actions that might harm others or are too selfish.

People should **pay attention** when the situation is dangerous, or the stakes are high, or others might suffer harm.

People should **monitor and pay sustained attention to** young children in their care.

People should not **plan or intend** to harm others without good reason (i.e., you or your society's flourishing and proliferation are threatened to a high enough degree).

People should **pay attention** to the effects and possible effects of their actions on others.

Planning, monitoring, and inhibiting are abilities both of persons and, at another level of analysis, of their brains. A claim such as "Cara is attending to the spider" has truth conditions at two levels, behavioral and mental. At the behavioral level, Cara must exhibit the behavior associated with attention, i.e., standing near enough to the tree, directing her gaze toward the spider, pausing for long enough. At the mental level, Cara's attentional processes must be operating in the right way. According to this approach, the truth conditions for mental state attributions, including the state of being culpable or responsible, exist at two levels. We assert a duplex proposition, with two parallel levels. Two states of affairs are "referred" to. When we assert the social functions or norms in that list, we are making an assertion about both the level of the person and the level of the mind/brain.[4]

For example, in addition to making a claim about people, the first truism—people should plan events, especially when they involve other people—also asserts that the executive process of planning should function in such a way that people are able to plan events, especially when they involve other people. Similarly, the second truism—people should stop or inhibit themselves from performing contemplated actions that might harm others, or are too selfish—also asserts that the process of inhibition should function in such a way that people don't do selfish or harmful things. What is different between these cases and other sorts of cases of biological functions such as pumping blood is that paying attention, planning, and inhibiting are things that *we* do, that typically reflect our values and desires. They exist at the level of persons (and at the level of their brains).

Fulfilling social functions and satisfying social norms requires that people behave in certain ways, and that parts of their minds behave in certain ways. Someone could satisfy every behavioral criterion for paying attention, for example, but fail to be paying attention because their mind is wandering. The mirror-image case, where a person is displaying no behavioral cues of paying attention to someone, yet is actually paying rapt attention to them, is quite possible. The behavioral side is metaphysically unnecessary, but often epistemically necessary.

The skeptics about the relevance of neuroscience to normative issues will surely say that "oughts" and "shoulds" apply only to full persons, and therefore there are no "oughts" or "shoulds" that apply to parts of persons. Yet the truisms appear to be cases in which "should" is (also) applied to a brain process. This implies that it is false that only persons have normative functions, or "shoulds." Brain processes can also have them. It would also be false that the ethical or normative realms emerge only at the level of the full person. There are ethical parts of us—the executive

[4] This responds to Patterson's concern that we are committing what Bennett and Hacker (2003) call the "mereological fallacy," which involves claiming that the referents of terms like "responsible" (and in fact all mental terms, such as "think," "see," or "attend to") can only be full persons, and specifically cannot be brains or brain parts or processes. Bennett and Hacker, and Wittgenstein, are correct in what they say about how we refer to behaving persons when ascribing mental states. But at the same time, other brain processes are representing the minds of those persons. Theories that do not take the second level into account are unable to explain certain linguistic phenomena, such as referential opacities.

processes—and they can do good or bad things, like persons can. They can do things that make us culpable.

Claims about the executive processes are appropriate objects of "should" clauses, because of their connection to social functions. Notice that these "shoulds" do not apply to other mental faculties. We cannot say that the color-blind person should see red as red and not green, except in a sense of "should" referring to the functions of the human visual system, in the way that the heart should pump blood. But we can say that parents should monitor their children at the beach and call them in if they are too far from shore. Executive processes are parts and processes in the world that we can apply these "shoulds" to, via duplex reference.

At the behavioral level the terms and predicates in the above truisms refer to people and the objects they interact with. At the mental level, what we can say is that our brains contain representations of the target person's mental states and processes. So as not to beg the question of whether what happens at the mental level truly counts as reference, I will call it "sub-reference." Using sub-reference, along with normal reference, is a way that the brain can employ two different representation systems to encode different types of features of an object. We can then speak about those features simultaneously, employing both systems.[5] In order to refer or sub-refer to x, one must possess an adequate concept, or more broadly, a representation of x. In the case of sub-reference we speculated in *RB* that there is a representational system tracking the referents and their properties: the default mode network.

### 4.3 Full "Shoulds"

Looked at one way, it seems obvious that science can inform us about what we should do. We can make inferences about what we should eat from the study of our digestive systems, for instance. But, are we sneaking in a normative element by assuming that the primary goal of eating is to keep us healthy, so that we can flourish, and proliferate our DNA? As I noted above, though, we need not look to the future to explain and defend claims about what we should eat. We can look to our evolutionary past and see that the parts of our digestive systems evolved to function in certain ways.

"People should eat a variety of foods," is a remark about how the human digestive system functions. As such, it is true. The "shoulds" of nutrition can carry normative weight, but not in the direct, forceful sense of the "shoulds" of the social functions. What then are the criteria for a brain process to have functions with full normative import, or full "shoulds"? It must have the right connections to our decision-making and control processes. In addition, it must have the right connections to brain systems responsible for our sense of self or identity. Here, the fact that we (sub) refer

[5] Rather than following Crimmens and Perry (1989) and referring to this phenomenon as "tacit reference," as we did in *RB*, I have switched to a more neutral term "sub-reference." Speaking of tacit reference might be taken to imply that there is no word token in the sentence that is doing the (sub) referring, which is often not true. Crimmens and Perry also do not hypothesize that two full propositions are expressed by the sorts of mental claims made above.

to executive activity with "I"—as in, "I planned the murder,"—is crucial because it signals a type of ownership of that activity. Another component necessary for a full "should" is the appropriate set of connections between the person's rationality and their identity. When the cognitive control network is in decision-making mode, those decisions are *my* decisions, for which I can be held responsible.

The condemnatory force emphasized by expressivist ethical theories comes partly from the knowledge that the target of the condemnation had conscious access to, e.g., the obvious dangerousness of a situation. There is also a connection between a person's sense of identity and the expressivism present in our ethical claims. The condemnatory force is augmented by the knowledge that the person made that decision in accord with their values and sense of self or identity. The force of the condemnation increases with each incident, because it becomes increasingly clear that this is *who* that person is. Sometimes we are also trying to teach a person with our full "should" claims. We are appealing to their emotions, in addition to their reason, to try to get them to attend to the situation and perhaps correct their error or change their future behavior.

### 4.4 The Setting of Moral Standards and Norms

Patterson objects that having a minimal working set of executive processes perhaps enables one to act responsibly, but that the executive processes are "in no way the measure of responsible action," and do not "set standards of responsible action" (this issue). There is a way in which this is true, since the executive processes help us hew to certain social norms and functions, and it is those norms and functions that constitute the current criteria for responsible action. But there is also a way, perhaps not the way that Patterson had in mind, in which the executive processes do set standards for responsible action. Social functions and norms are set via our evolutionary course, the environments we found ourselves in, and the relation between our society and others nearby, along with the structures of our bodies and brains and their functional systems. Having executive processes allowed us to form much more complex and cohesive societies, by getting people to adopt and follow a complex system of norms. We humans live up to standards that lions and tigers cannot live up to, and that is why we do not hold them responsible for killing humans. The reason why we can live up to these higher standards is because our brains have executive processes and theirs do not. We can sustain our attention for much longer than they can. We can plan with much greater detail and further into the future than they can. We can inhibit ourselves far better than they can. We can make much more informed and logical decisions than they can. Human societies become possible only with these advanced abilities, and our functions and norms assume them and rely on them and must themselves hew to them in many ways.

Being a proper functioning member of a society requires lots of executive activity. Societies are complex and require all our cognitive resources to negotiate. Our ability to fulfill social functions and hew to norms depends heavily on having executive processes. They enable high-level correction of our actions based on our situation. Minding children, for instance, requires deliberately sustained attention for hours

at a time. Our social functions are more complicated, must be pursued over longer periods of time, and bind us more strongly than the social functions of any non-human social animal, such as canines, felines, or other primates. Of course, many people naturally and unreflectively abide by the functions and norms and behave in ways that support them. Many parents reflexively care for their children. But the role of executive processes is to *make sure* that parents care for their children.

Our executive processes and the roles they play in allowing us to be responsible would have played a large part in the creation of our concept of responsibility. It is tailored to them in many fundamental ways. It encapsulates information about the subset of our behaviors we can be held responsible for, and the subset of our behaviors we can be blamed and punished for. Over time, the executive processes played crucial causal roles in generating the standards we have. At some point in our evolution, we moved beyond the other social animals and developed a normative aspect—we became proper subjects of full "shoulds."

## 5 Objection: Neuroscience is Wholly Dependent on the Connections it has to Psychology and Behavior for any Relevance it has to Responsibility

"To the extent neuroscience can be useful," says Stephen Morse, "it is virtually entirely dependent on well-validated psychology to correlate with the neuroscientific variables under investigation" (2021, this issue). To support this, he argues that, "neuroscientists do not go on expensive fishing expeditions without knowing what they are hoping to catch. Instead, they have already identified some psychological or behavioral trait or condition, such as impulsivity, addiction or schizophrenia, that already interests them theoretically or practically" (*Ibid.*).

Very often, though, important discoveries are made not by working from the top down, using our knowledge of psychology and behavior, but from the bottom up, using knowledge about brains. In the early 1990s, cognitive neuroscientist Antonio Damasio treated a patient named Elliot who had developed a large tumor in his orbitofrontal cortex, the part of the cortex just above the eye sockets. Elliot's friends became aware that he had a problem, when decisions kept turning out badly for him, unlike his generally prudent decisions prior to the tumor. But his doctors couldn't figure out what the problem was, and they could not characterize it in behavioral terms. They needed to know what had gone wrong with his decision-making process.

A fishing expedition is exactly what Damasio undertook. He knew that his patient had a rather large brain lesion in the orbitofrontal cortex that must have disabled some function related to decision-making. He gave his patient every mental test at his disposal, which Elliot scored well on. Damasio also looked at the connections that the damaged area had to other brain systems and found that it had strong connections to the autonomic system. In the end, he discovered that there is a part of the autonomic system that plays crucial roles in our decision-making and that, in his patient, components of the ability to make a gut-level, prima facie assessment of the viability and riskiness of plans were damaged. Working from this, he was able with

his colleagues to construct a new behavioral test for Elliot's deficit, the Iowa Gambling Task (Bechara et al. 2000).

Another case of going from the neuroscience to the behavior: As part of the rehabilitation process, people recovering from brain damage are tested by a neurologist to assess their brain function. Neurologists can quickly test different domains of mental activity using their large toolset of tasks and stimuli. But based on what they know about the part of the brain that was damaged, they will have reason to believe that certain functions are compromised, prior to seeing any behavior from the patient, and this allows them to focus their examination.

Another case: Scientists studying fine motor activity of the hands in chimpanzees found that the areas they were recording from also became active whenever their monkeys *saw* someone pick up something, or eat something (Rizzolatti 2004). This was a revelation to them because they assumed these cells functioned only to control motor activity of the hands. But it turned out that their brains, and ours, use these hand representations to perceive the hands of others. They had discovered mirror neurons. This has taught us several things about how we perceive the actions of others or, in some conditions, fail to do so.

Morse says that the presence or not of a minimal working set of executive processes must be determined behaviorally. But we can determine that a person lacks a minimal working set of executive processes using neuroscience alone, in cases where we have no behavioral evidence at all. Consider the example of a person who has just had a stroke that has damaged a certain portion of his orbitofrontal cortex, which will produce disinhibited and possibly dangerous behavior, yet who has not yet moved a muscle. We know at this point that he is less responsible for any subsequent harmful behavior, since he has lost a large degree of control over it. We have the technology to do this with a fairly high degree of certainty now, and that degree will only increase. Neuroscience can tell us not only that, but why, a person has or lacks a minimal working set of executive processes, without appeal to behavior.

### 5.1 Separating Metaphysics from Epistemology

Our legal system respects the epistemic/metaphysical distinction at several vital places. For example, our ability to know which of several suspects actually committed a crime is often very limited. Yet we still need to speak of the actual perpetrator, even if we never find that person. Our ability to discern guilt and innocence is often quite limited, yet we still need to speak of the actual guilt or innocence of defendants or suspects. And we always need to mind the distinction between the jury's determination of guilt or innocence and what is in fact the case.

In the legal world the distinction between the metaphysics of culpability and its epistemology corresponds to the distinction between the mental components of the crime itself and how we learn about them. *Mens rea* is part of the crime itself. Legal doctrines refer to mental states and processes, which they assume are hidden inside the minds of people. Only those people can be certain about their occurrence and nature; everyone else must make inferences, according to these assumptions. This is also the basic metaphysical approach that folk psychology adopts.

In seeking to learn more about the crime itself, we need to be clear about what our target is. In particular, we need to distinguish between the crime and how we acquire evidence of the crime. The crime itself—the overt acts and other public events—as is the case with any event or state, can be known about in multiple ways. There are many epistemic routes to the crime, and there are many relevant events, including behaviors, that could be observed via these routes. The legal system already has numerous ways to gain knowledge of crimes: eyewitness testimony, audio and video recordings, blood typing, fingerprinting, various types of documents, police/penal records, and medical reports such as x-rays or accounts of hospital visits. The mental part of the crime can also be known about in multiple ways, e.g., via behavior, via neurologists' reports, via EEG, via brain imaging, and so on. What these latter ways have in common, what unifies them and allows them to count as evidence, is their causal connections to the mental events at the heart of the crime—such as planning, forming intentions, monitoring, and task switching—directed toward criminal activity, all of which, assuming physicalism, are brain processes.

"*Mens rea*" does not refer to behavior, or to behavioral dispositions. Behavior is evidence of *mens rea*. Morse's approach gets the metaphysics wrong. "*Mens rea"* refers to events hidden inside the skull of the defendant, events that played certain crucial causal roles in producing criminal behavior or behavior relevant to their guilt. No matter how remote, indirect, improbable, or seemingly impossible our contact is with the phenomenon itself, we must mind the distinction between what something is and how we know about it. If we do, we have in neuroscience a theory with an ontology that is capable of aligning with the ontologies of both folk psychology and legal thought, in postulating inner mental events, which neuroscience conceptualizes as brain events.

Morse is right that we often use our knowledge of psychology to *interpret* results in neuroscience, but the converse also holds: neuroscience can help us understand puzzling psychological phenomena, as we saw above. In this way, the relationship between the two fields is symmetrical, and their methods are complementary. Of course in writing *RB* we already had a concept of responsible action. But it was vastly expanded, corrected, and informed by our study of the neuroscience. Sometimes we use our knowledge of psychology to identify the neuroscientific phenomenon of interest. But a clear distinction can be made between how we identify something, and what that thing itself is. We identify water as the colorless, tasteless, clear liquid that we require for life. But water itself is $H_2O$. Once the identification is made, we reconceptualize water as $H_2O$, and we have a new way to identify water, and a science to study it with.

The criteria for responsibility should refer to the actual phenomena that play the crucial causal roles. We consider it a virtue of our account that it honors the distinction between the mental events that make us culpable, and criminal behavior, just as folk psychology and the law do. Executive processes are the core *mentes reae*, on our account. Their activities make the person culpable. Observable behavior such as saying, "I intend to kill x," is defeasible, in the sense that one can be kidding, drunk, not competent, etc. But the sane, rational intent to kill, present in the mind of the perpetrator, is the law's focus.

In response to our claim that a collection of tests might in the future be able to accurately determine the presence of a minimal working set to a degree where it could be used in legal cases, Morse argues that "such tests would depend on first identifying, well-characterizing and operationalizing measures of the behavioral deficiencies that might compromise the presence of a MWS" (this issue). Notice the oddness of saying that behavioral deficiencies might compromise the presence of a MWS. It goes the other way around: Failure to have a MWS causes behavioral deficiencies. The inability of a car to stop does not compromise its brakes, the malfunctioning brakes compromise its ability to stop.

There are some other undesirable consequences of attempting to use behavior in the way that Morse does. There are behaviorally indiscernible cases with vastly different levels of culpability: Two people are admiring a view from a great height. One appears to trip, bumps into the other, and over that person goes. Unfortunate accident or clever murder? Only knowledge of *mens rea* can help us make that determination. Another case: What a person meant by certain words they said is often crucial in legal cases. But speech behavior underdetermines mental states, such as the state of meaning this or that. The reason why Quine's gavagai paradox (Quine 1960) arises is that we are only allowing ourselves to look at the behavior of the natives (Searle 1987). This produces an indeterminacy of reference, where "gavagai" (said by a hunter after a rabbit appears) could have one of a large number of distinct meanings. But there is a fact of the matter as to what a person means by something they say, that we are aware of in our own cases. This awareness can eventually be understood and described by neuroscience.

Lastly, the reliance on behaviorism misses the fact that we have other epistemic resources for learning about the minds of others than via observing behavior. We have mindreading capacities, one of which is the default mode network, which is working to represent the minds of the people we make responsibility claims about, including claims about what they should have done or not done. This is an important omission. If our grasp of *mens rea* occurs substantially via a mindreading or theory of mind ability such as the one enabled by the default mode network (Buckner et al. 2008; Li et al. 2014), this means that it involves a non-behavioristic epistemic route to *mens rea*.

### 5.2 Schizophrenia

Morse refers to schizophrenia as a "psychological or behavioral trait," and says that, "one cannot study schizophrenia neuroscientifically, for example, until behavioral criteria for the condition have already been developed using clear cases" (this issue). But this is not how the DSM (APA 2013) states it. According to the current version, the DSM-5, there are five criteria for a diagnosis of schizophrenia. There is a core set of three, one of which must be present. These are (1) delusions, (2) hallucinations, and (3) disorganized speech. In addition, two other criteria may be used: (4) disorganized or catatonic behavior, and (5) "negative" symptoms, such as reduced emotional expressiveness, decreased motivation, and decreased interest in social

activities. Note that the first two of the three core symptoms—delusions and hallucinations—describe inner mental events or states. Notice also that the presumption is that the external symptoms such as disorganized speech and reduced emotional expression are not being produced deliberately, but are rather a sign of some problem(s) with the person's underlying cognitive or emotional systems.

Scientists—in this case the producers and consumers of the DSM—are acting exactly as the law does in making essential reference to mental events that are in fact brain events. In its criteria for schizophrenia, the DSM lists inner mental events and states, such as hallucinations, delusions, and emotions. And in its criteria for responsibility, the legal system lists inner mental events such as planning and intending—the events that can constitute a guilty mind. Since psychology moved away from behaviorism decades ago, toward cognitive neuropsychology, a brain-based ontology is a better fit for it now.

## 6 Objection: There are Problems with the Analysis of the Kenneth Parks Case in *Responsible Brains*

What is the culpability of a person who harms others *while he is asleep*? Sleepwalking is a well-known occurrence that affects 3.6% of the population (Ohayon et al. 2012). It occurs during non-REM sleep. Typically, sleepwalkers get up and perform routine actions, such as eating food from the cupboards or refrigerator. The actions can be fairly complicated, such as making a passable sandwich, they just need to be well-practiced, or routinized. REM behavior disorder (REMBD) is a rarer condition, found among 1% of the population (Haba-Rubio et al. 2018). It is associated with neurodegenerative disorders such as Parkinson's disease and Lewy body dementia (Boeve 2010). Unlike sleepwalkers, those with REMBD are active during REM sleep. In general, the existence of REM sleep correlates well with the existence of dreaming. Their actions are driven by the dream events, but adapted to their actual surroundings. For example, a woman dreamed that her house was on fire, and that the fire crew was yelling at her to throw her baby out the window. She, alas, threw her real baby out of a real window. Because of their dream-based genesis, the actions of people with REMBD are not routine, and can be quite complicated and intelligent-looking, as well as dangerous to the dreamer and others.

Morse provides a nicely detailed analysis of the Parks' case that I will respond to point by point. Morse treats Parks as a sleepwalker, but Parks fits the diagnosis of REMBD better, I would argue, since what he did was not a routine action and showed some level of behavioral flexibility. If so, then his variant would be categorized as idiopathic REMBD, as opposed to the more common form of REMBD caused by a neurodegenerative disorder.

The difference between sleepwalking and REM behavior is crucial for properly assessing Parks' level of culpability, I will argue. Morse says that in sleepwalking, "the agent is clearly responsive to environmental cues and is seemingly goal-directed" (this issue). But this applies to both sleepwalkers and those with REMBD. What differentiates them behavior-wise is the ability of those with REMBD to

engage in non-stereotyped behaviors, which are more complex and flexible. "Some sleepwalkers raid the refrigerator and some commit homicide," (*Ibid.*) says Morse. Raiding the fridge is a routine, stereotypical behavior of sleepwalkers. Homicide is not a routine behavior and requires the flexibility one sees in REMBD.

Morse says that "[d]espite his sleepwalking state, Parks' behavior exhibited many aspects of executive functioning. He attended to his environment and carried out a number of intricate instrumental tasks to accomplish his homicidal result" (this issue). This is also true of the midnight sandwich maker, who is performing a highly routinized action without executive involvement. What differentiates the two is that the ability to go beyond preprogrammed behaviors are greater in REMBD. There is good evidence that the executive processes are offline during dreaming and REM sleep (*RB*, p. 143), which would explain why we are sometimes irrational in our dreams. This fact can also explain the irrationality of the actions of people with REMBD, and it can ultimately absolve them of responsibility (we argued in *RB* that it absolves Parks).

According to Morse, sleepwalkers lack "a crucial component of rational self-regulation, which is self-monitoring," (this issue) as well as rational capacity. Indeed, but self-monitoring is an executive process. It is effortful, it involves consciousness and representations, and it involves reacting intelligently and in a non-stereotypical way to the environment, for instance by detecting and correcting for problems that arise during the commission of an act. Referring to Parks, Morse says that "[t]hrough no fault of his own, he had lost the ability to monitor himself, the ability we all fundamentally use to guide our own behavior. In a sense, he wasn't watching himself when he did his horrible deeds and thus could not bring good reason to bear" (*Ibid.*). We could offer no better example of sub-reference to executive processes. The sub-reference of, "he wasn't watching himself," is that his cognitive control network was not causally interacting with the appropriate brain processes (these would include his visual system, since he is seeing himself hurt his in-laws). By building a theory stated in frank neuroscientific claims, we can move beyond the metaphor of watching oneself to assign concrete referents to that claim.

There are also problems with using the ability to reason alone as a criterion for eligibility for culpability. Reasoning depends on a sound executive, but not vice versa. There are cases that are excused that do not involve inability to reason at the time. But they do involve executive problems—disinhibition, for example. In addition, being unable to reason at the time seems to exclude too much, such as road-rage crimes and acts driven by powerful jealousy. So, being unable to reason is both too broad and too narrow as an excusing condition by itself. We agree that Parks cannot reason, but we can say why: His executive processes are offline, as they typically are during dreams.

Morse: "*RB* emphasizes that Parks was not 'there,' but this seems incorrect. Even if Parks fully believes that he should have been rightfully acquitted, he still must acknowledge to himself that he did kill people rather than raid the refrigerator" (this issue). I would argue that this is a sense of "he" that can also be present in accidental killings. It is still true to say of a faultless driver that he killed the person who suddenly bolted in front of his car. In saying that Parks was not present, and that the actions are not his, we are emphasizing that his executive processes are inactivated.

Parks' actions that night did not have "executive approval." One thing this approval accomplishes is to ensure that contemplated actions are in the service of one's own desires and values, and there is no evidence that Parks himself desired to harm his in-laws.

REM behavior disorder carries less culpability than sleepwalking because REMBD can produce non-routine behavior. Non-routine behavior has the potential to be much more dangerous than routine behavior. If Parks had been shown to be subject to sleepwalking and not REMBD, his culpability would be much greater, since his non-routine actions speak in favor of a higher degree of awareness and "presentness" not covered by that diagnosis. This assumes that it is the first time the sleeper has harmed people while in an episode of REMBD, which was true in the case of Parks. But if harm has happened before, the REMBD sufferer may be guilty of negligence, for failing to take steps to prevent it.

## 7 Objection: *Responsible Brains* Owes us an Account of Abilities

It will help in this discussion to distinguish between mental capacities and behavioral capacities, and to try to get clear on how each are connected to the executive processes. The cognitive control network gives humans certain mental abilities, and this enables certain behavioral abilities or capacities. In contrast, Morse refers to the executive processes themselves as "behavioral capacities." If one thinks of executive functions as a set of external abilities, or actions, or behaviors, neuroscience might indeed seem redundant. They are not behavioral capacities unless we are speaking about the behavior of a brain network. In this section I will respond to Michael Moore's penetrating critique of what sort of capacities or abilities, which he correctly notes are mental, it is that the executive processes endow us with.

Moore (2021) cites a case from Rangel et al. (2009) that refers to a finding in which subjects showed activity in the dorsolateral prefrontal cortex (DSLPFC) when they successfully resisted an urge to eat cake. "In a case where the subject eats the cake and his [ventromedial prefrontal cortex] is active but his DSLPFC is not," Moore asks, "is the subject responsible for going off his diet?" (this issue). Assuming the scientists have found the right spot, the activity of the DSLPFC *is* the subject's act of resisting the cake, or inhibiting the impulse to eat cake from developing into an intention to grab a piece. The subject is responsible because the DSLPFC should have activated, and because we (correctly) accept ownership and responsibility for what the cognitive control network does (assuming we are speaking about a portion of the DSLPFC that is part of the cognitive control network). That is, this is also partly a question about the relation the DSLPFC has to one's self and identity. If we assume that, along with the right connections to that person's identity or sense of self, the DSLPFC is doing the appropriate executive work, its failure to activate makes the subject culpable.

There is still a question, though, about exactly how the person's abilities are grounded in the abilities/capacities of the cognitive control network. Moore supposes that our approach holds that "the (dis)abilities of various brain structures translate directly into the (dis)abilities of whole persons that are relevant to

responsibility" (*Ibid*.). Yes, according to the executive account, disabilities of brain structures do translate straightforwardly into disabilities of persons. Thanks to the phenomenon of duplex propositions (in which we refer simultaneously to the person and sub-refer to her executive processes), we can bind the person and her cognitive control network together as the referents of a single claim. A malfunctioning executive process of planning makes the person unable to plan. A malfunctioning process of inhibition makes them unable to inhibit.

Care is needed in describing how executive activity relates to the person as a whole. When we say that a person did or did not activate his executive processes, this simply means that his executive processes did or did not activate, in most cases. But, in some cases, there is a difference between, e.g., "Jan's process of inhibition activated," and, "Jan inhibited (or stopped) herself." In obsessive–compulsive disorder, for example, the process of inhibition repeatedly activates in an unwanted and disowned way as being, e.g., inconsistent with one's values and desires.

Moore notes, however, that there are distinct differences in the causal powers of a person and a cognitive control network. He points out that "different abilities (say of the DSPFC and of the person whose DSPFC this is) are analyzed by different counterfactuals … so there is no reason to think that necessarily if a person's DSPFC lacks an ability to modulate signals from the vmPFC then the person must be unable to choose other than they did" (this issue). The causal powers of the cognitive control network and the person who owns it are different, and this implies that different sets of counterfactuals will apply to the two, as Moore notes. But the counterfactuals will be similar in important respects. Our mental powers correspond to the causal powers of the executive processes at many points. For example, we can alter the way we respond to a stimulus, but we cannot alter the way that stimulus appears to us (e.g., see it as red when it is blue). In the places where the two sets of counterfactuals fail to overlap lie all sorts of interesting cases. For instance, some of these cases will involve our falsely thinking we have control when actually the cognitive control network does not have control. Ultimately, since we are talking about two physical systems, one of which "contains" the other, the question about how the two sets of counterfactuals relate is the same as the question in the philosophy of science about how different levels of analysis, such as physics and chemistry, relate to one another. For example, how do the causal powers of atoms relate to the causal powers of molecules?

## 8 Objection: The Account of Negligence in *Responsible Brains* is Flawed

Craig Agule (2021, this issue) suggests an interesting simplification of our criterion 3, to eliminate the need for stating it "negatively," i.e., speaking about what should have happened but did not. He proposes the following change to our three criteria: 1. they have a MWS, 2. they act or omit to act, and 3. the person's MWS plays a role in the agent's action or omission.

I don't believe the changes to criterion 3 improve it, however. The examples Agule gives of executive activity playing a "positive" role in the omission cases are

problematic. They involve routine behavior without real executive intervention such as making a sandwich or driving. Of course, these two activities are not always done in routine mode. One might well behave the same way, with plenty of executive involvement, if one had never made a sandwich before, or one was a sandwich chef, experimenting with new ingredients. Likewise, driving can be done on autopilot, or with intense attention, thought, and planning. But even in these positive cases, the executive activity that is present is very often not relevant to culpability.[6] Rather, we can look into the person's past and find specific occasions when they should have employed executive processing but did not.

Agule's Sam and Ruth case can also be analyzed "negatively," as a failure to employ the right executive processing. Sam and Ruth are parents who get distracted while hosting a dinner party and forget to tend to a child in the bath, who drowns. They failed to *sustain* their attention for a pre-set period. Depending on how long attention must be sustained, different techniques are available for making sure one gets it right. I can make a mental note that I am boiling water, I can use an egg timer, or I can use the alarm function on my phone. The criteria for sustaining attention in this case were clear: the time it takes a bathtub to fill. Sam and Ruth's fault is a monitoring failure. Monitoring is a type of sustained attention, which involves a sense of mental effort, as would be expected with an executive process.

In negligence cases, we can say that the appropriate executive activity did not happen, but should have. The more obvious it was that executive activity needed to take place, the higher the culpability, and the greater the negligence, yielding this formula for determining negligence and its degree of severity: Negligence occurs when executive activity did not happen, but should have. (The "should have" clause of 3 is used, and criteria 1 and 2 are satisfied). The degree of severity depends on: a) how many times this has happened in the past; b) how obvious (clear, apparent, manifest) it was that the situation required executive activity—the more obvious the danger, the greater the culpability; c) whether the executive activity that did not occur had a reasonable chance of preventing something bad from happening—the greater the chance, the greater the culpability and the severity of the offense; and d) how bad the consequences were; the worse the consequences, the greater the severity of the negligence. As applied to Sam and Ruth: a) we need to know if they have been neglectful or negligent with their children in the past, as this can have a big effect on our judgments; b) we need to know how obvious it would have been to them that they should check on their child—if a friend reminded them, or they heard noises coming from the bathroom but didn't check, their culpability rises; c) certainly the odds of the right sort of executive activity—such as sustained attention—preventing the harm were very high; and d) the consequences were momentous.

[6] There may be a conceptual problem with Agule's version of criterion 3, since the concept of an omission seems very close to the concept of a failure, which I argued above contains a "should have.".

## 9 Conclusion

If the discussions above are relevant to responsibility, that is because neuroscience is relevant to responsibility. I have also argued in several different ways that neuroscience is relevant to normative issues in general. Neuroscience-based approaches can help us make more accurate and fine-grained assessments of culpability and its severity. They can also allow us to understand how our minds, using executive processes, can guide us to conform to a complex set of social functions and norms. But those functions and norms also must conform to the executive processes themselves.

**Acknowledgements** I am indebted to the commentators for their focused and edifying remarks. Special thanks to Dennis Patterson for working to set up our Covid-cancelled conference, then assisting in the production of this issue. Thanks to Melinda Campbell, Andreas Kuersten, and Katrina Sifferd for comments.

**Authors contributions** Solo authored.

**Funding** Prof. Hirstein's work was funded by a Templeton grant on The Philosophy and Science of Self-Control, administered by Alfred Mele.

**Declarations**

**Conflict of interest** There are no conflicts of interest.

## References

American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, Fifth edition (Arlington, VA: American Psychiatric Association, 2013).

Agule, C. (2021). "Minding Negligence", *Criminal Law and Philosophy*.

Baars, B. (1997). *In the Theater of Consciousness*. New York: Oxford University Press.

Bechara, A., Damasio, H., Damasio, A. (2000). "Emotion, Decision Making, and the Orbitofrontal Cortex", *Cerebral Cortex* 10(3): 295–307.

Bennett, M.R., Hacker, P.M.S. (2003). *Philosophical Foundations of Neuroscience*. Hoboken: Wiley.

Berker, S. (2009). "The Normative Insignificance of Neuroscience", *Philosophy and Public Affairs* 37(4): 293–329.

Block, N. (1995). "On a Confusion about the Role of Consciousness", *The Behavioral and Brain Sciences* 18: 227–287.

Boeve, B.F. (2010). "REM Sleep Behavior Disorder: Updated Review of the Core Features, the RBD-Neurodegenerative Disease Association, Evolving Concepts, Controversies, and Future Directions", *Annals of the New York Academy of Sciences* 1184: 15–54.

Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L. (2008). "The Brain's Default Network: Anatomy, Function, and Relevance to Disease", *Annals of the New York Academy of Sciences* 1124: 1–38.

Crimmens, M., Perry, J. (1989). "The Prince and the Phone Booth: Reporting Puzzling Beliefs", *The Journal of Philosophy* 86(12): 685–711.

Crone, E.A., van der Molen, M.W. (2004). "Developmental Changes in Real-Life Decision Making: Performance on a Gambling Task Previously Shown to Depend on the Ventromedial Prefrontal Cortex", *Developmental Neuropsychology* 25(3): 251–279.

Coppola, F. (2021). "We are More Than Our Executive Functions: On the Emotional and Situational Aspects of Criminal Responsibility and Punishment", *Criminal Law and Philosophy*.

Doris, J. (2017). *Talking to Ourselves: Reflection, Ignorance, and Agency*. New York: Oxford University Press.

Eickhoff, S.B., Todd Constable, R., Thomas Yeo, B.T. (2018). Topographical Organization of the Cerebral Cortex and Brain Cartography, *Neuroimage*, 170: 332–347.
Fischer, M.J., Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.
Fan, S., van den Heuvel, O.A., Cath, D.C., de Wit, S.J., Vriend, C., Veltman, D.J., ven der Werf, Y.D. (2018). "Altered Functional Connectivity in Resting State Networks in Tourette's Disorder", *Frontiers in Human Neuroscience* 18: online.
Husak, D., (2021). The Objective(s) of Responsible Brains". *Criminal Law and Philosophy*.
Hampson, M., Tokoglu, F., King, R.A., Todd Constable, R., Leckman, J.F. (2009). "Brain Areas Coactivating with Motor Cortex During Chronic Motor Tics and Intentional Movements", *Biological Psychiatry* 65: 594–599.
Hirstein, W. (2012). *Mindmelding: Consciousness, Neuroscience and the Mind's Privacy*. New York: Oxford University Press.
Hirstein, W., Sifferd, K.L., Fagan, T.K. (2018). *Responsible Brains: Neuroscience, Law, and Human Culpability*. Cambridge, MA: The MIT Press.
Haba-Rubio, J., Frauscher, B., Marques-Vidal, P., Toriel, J., Tobback, N, Andries, D., Preisig, M., Vollenweider, P., Postuma, R., Heinzer, R. (2018). Prevalence and Determinants of Rapid Eye Movement Sleep Behavior Disorder in the General Population", *Sleep* 41(2): online.
Levy, N. (2014). *Consciousness and Moral Responsibility*. New York: Oxford University Press.
Li, W., Mai, X., Liu, C. (2014). "The Default Mode Network and Social Understanding of Others: What Do Brain Connectivity Studies Tell Us?", *Frontiers in Human Neuroscience*. Online: https://doi.org/10.3389/fnhum.2014.00074
Libet, B. (1985). "Unconscious Cerebral Initiative and the Role of Consciousness in Voluntary Action", *The Behavioral and Brain Sciences* 8: 529–539.
Milliken, R. (1989). "In Defense of Proper Function", *Philosophy of Science* 56: 288–302.
Moore, M.S. (2021) "Relating Neuroscience to Responsibility: Comments on Hirstein, Sifferd, and Fagan's Responsible Brains", *Criminal Law and Philosophy*.
Morse, S. (2021). "Is Executive Function the Universal Acid?". *Criminal Law and Philosophy*.
Neander, K. (1991). "Functions as Selected Effects: The Conceptual Analyst's Defense", *Philosophy of Science* 58: 168–184.
Ohayon, M.M., Mahowald, M.W., Dauvilliers, Y, Krystal, A.D., Leger, D. (2012). "Prevalence and Comorbidity of Nocturnal Wandering in the US Adult General Population", *Neurology* 78(20): 1583–1589.
Patterson, D. (2021). "Inert". Criminal Law and Philosophy.
Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: The MIT Press.
Rangel, A., et al. (2009). [cited in Moore's response, could not find]
Rizzolatti, G. (2004). "The Mirror Neuron System", *Annual Review of Neuroscience* 27: 169–192.
Searle, J.R. (1987). "Indeterminacy, Empiricism, and the First Person", *The Journal of Philosophy* 84(7): 123–146.
Spencer H. (1860). "The Social Organism", *Westminster Review* 73: 51–68.
Sripada, C. (2016). "Self-Expression: A Deep Self Theory of Moral Responsibility", *Philosophical Studies* 173(5): 1203–1232.
Wright, L. (1973). "Functions", *Philosophical Review* 82: 139–168.
Yeo, B.T.T., Kreinen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zollei, L., Polimeni, J. R., Fischl, B., Liu, H., and Buckner, R. L. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *The Journal of Neurophysiology* 106(3): 1125-65.