

THE UNIVERSITY OF
ALABAMA

SCHOOL OF LAW

Minds, Brains, and Norms

Michael S. Pardo

Dennis Patterson

Neuroethics, forthcoming

This paper can be downloaded without charge from the Social
Science Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=1432476>

Minds, Brains, and Norms

Michael S. Pardo* and Dennis Patterson**

Debating the philosophical merits of a neuroscientific approach to matters of mind requires, at a minimum, a clear articulation both of the subject matter of the debate (i.e., the disputed questions) and some idea of what success or failure in argument might look like. In this article, we start with a statement of some of the things we do not contest. We then move on to articulate a conception of philosophy, one we think gives us a critical stance. Finally, we address three areas in which scholars have claimed that neuroscience can inform matters of ethics, value, and law. These areas include (1) unconscious rule following, (2) interpretation, and (3) knowledge. With regard to these three matters, we register our dissent from the current orthodoxy in neuroethics.

The philosophical issues raised by the intersection of ethics and neuroscience traverse a wide range of both subjects and questions of philosophical interest. The topics include, for example, consciousness, free will, understanding and interpretation, and the nature of human action. In an earlier article on the interconnections between neuroscience and law,¹ we focused our attention on the broad question of the relationship between the brain and human action in the legal context. As a way of introducing the issues we focus upon here, we will make just a few remarks about our starting point vis-à-vis our interlocutors.

* Associate Professor, University of Alabama School of Law. Professor Pardo thanks Dean Ken Randall and the University of Alabama Law School Foundation for generous research support.

** Professor of Law, European University Institute, Florence; Board of Governors Professor of Law and Philosophy, Rutgers University School of Law, Camden; Professor of Jurisprudence and International Trade, Swansea, UK.

¹ Michael S. Pardo & Dennis Patterson, "Philosophical Foundations of Law and Neuroscience," *Illinois Law Review* (forthcoming 2010), available at <http://ssrn.com/abstract=1338763>.

We in no way take issue with the claim that the brain is intimately related with mental life.² We recognize that the feelings and activities that we associate with the mind depend upon a (properly functioning) brain. Particular neurological states, in others words, may be a necessary condition for various mental activities. Finally, we do not contest that neuroscience may illuminate how these activities depend upon the brain and how damage or defects in the brain may affect one’s mental activities, including whether one has the capacity to engage in these activities at all. In short, we are not dualists, and we do not object to scientific approaches to understanding the brain, mind, and human behavior, and the relationships between them. So where do our objections begin?

In his recent book *Out of Our Heads*,³ Alva Noe makes the point—one with which we agree—that the basic problem with the philosophical assumptions of much writing about neuroscience is that the picture of human action it produces is one in which the person is reduced to the brain. In short, the thesis is “you are your brain.”⁴ We share Noe’s view that the brain is part of who we are, and moreover, that understanding human behavior requires a look at what Noe calls the “whole animal.”⁵

² We agree completely with the view that “[t]he brain is necessary but not sufficient to account for all the physiological and psychological properties that make each of us a unique person.” Walter Glannon, “Our Brains are Not Us,” 23 *Bioethics* 321, 321 (2009).

³ Alva Noe, *Out of Our Heads* (2009).

⁴ For a sophisticated defense of this thesis, see Joshua Greene & Jonathan Cohen, “For the Law, Neuroscience Changes Nothing and Everything,” 359 *Phil. Tran. Royal Soc’y London* 1775 (2004)

⁵ *Id.* at 6 (“we need to look to the ways in which each of us, as a whole animal, carries on the processes of living in and with and in response to the world around us.”). See also Walter Glannon, “Our Brain are Not Us,” 23 *Bioethics* 321 (2009). Max Bennett and Peter Hacker have coined the phrase “the mereological fallacy” to refer to the error of ascribing properties to a part of an animal (the brain) that make sense only when ascribed to the animal as a whole. See Maxwell Bennett & P.M.S. Hacker, *Philosophical Foundations of Neuroscience* 68-74 (2003). In Pardo & Patterson, *supra* note 1, we discuss several examples of the mereological fallacy within the neuro-law literature.

This leads us to what we regard as the “big picture” difference between proponents of the reductive thesis that “you are your brain” and those who, like us, insist on a wider field of vision in the explanation of human action. In this regard, it is important to remember that what we are all after is the best characterization of human action—and, more generally, what it is to be “human.” We believe the idea that “you are your brain” simply leaves too much out of its picture of human action for the picture to be plausible.

The plausibility or implausibility of different accounts of human action shows up in the particulars of explanation. For example, many scholars writing on neuro-ethical issues believe that when it comes to human decision making, all the work is done in the brain.⁶ Tendentiously expressed, the view is that human decisions are literally “made” in a certain region of the brain. As such, the brain itself can be studied so that we might learn just how it is that humans decide on one course of action over another, weigh the consequences of a decision, and render judgments of appropriateness.

We said that the plausibility of explanations of human action show up in detailed accounts of how humans engage in activities of mind. For example, is it plausible to think that human decision making and human knowledge can be explained solely by reference to brain function? We think not. We believe that a single-minded focus on the brain leads to explanations that are implausible. In this article, we concentrate on these difficulties in an effort to reveal the explanatory weaknesses of this single-minded approach. To that end, we focus on the topics of rule following, interpretation, and knowledge.

⁶ See, e.g., Joshua Greene & Jonathan Haidt, “How (and Where) Does Moral Judgment Work?,” 6 *Trends in Cog. Sci.* 517 (2002); Terrence Chorvat & Kevin McCabe, “The Brain and the Law,” *Law & the Brain* 128 (Semir Zeki & Oliver Goodenough eds., 2006); Alan G. Sanfey et al., “Neuroeconomics: Cross-Currents in Research on Decision-Making,” 10 *Trends in Cog. Sci.* 108 (2006); Oliver R. Goodenough, “Mapping Cortical Areas Associated with Legal Reasoning and Moral Intuition,” 41 *Jurimetrics J.* 420 (2001).

What unites these three areas is that they each involve abilities engaged in by people as a whole in particular contexts.⁷ What unites our discussion is a particular methodological approach. Our concern is with the philosophical foundations underlying these areas—that is, with conceptual questions involving the proper application of the relevant concepts, not with empirical questions regarding the brain and its functions.⁸ Our focus is on the criteria that apply to ascriptions of successful examples of rule following, interpretation, and knowledge.⁹ As we will argue, these criteria involve behaviors of various kinds by people, not neurological states of their brains. The upshot of this conclusion is not that claims that the brain “follows rules,” “interprets,” and “knows” are false; it is that these claims are lacking in sense, at least to the extent these terms are being used in *their* ordinary sense.

I. Unconscious Rule Following

One of the most basic questions of ethics and law concerns norms and conformity (or lack of it) with them. Interest in this question stems from the desire to learn more about the nature of moral cognition; how it is that we decide what norms there are and what is required by those norms. This is the issue of norm application or, in the language of some philosophers, what it means to follow a rule.

Many scholars take the view that moral knowledge is “encoded” or “embedded” in the brain.¹⁰ This philosophical rationalist view of the nature of moral knowledge views the capacity

⁷ The areas are related in a deeper sense as well: to follow a rule or norm and to interpret a rule or norm both involve the application of one’s knowledge.

⁸ M.R. Bennett & P.M.S. Hacker, *A History of Cognitive Neuroscience* (2008).

⁹ The criteria establish the norms for ascriptions of these concepts.

¹⁰ See, e.g., John Mikhail, “Moral Grammar and Intuitive Jurisprudence: A Formal Model of Unconscious Moral and Legal Knowledge,” *50 Psy. of Learning and Motivation* 27,28 (2009) (“The moral grammar hypothesis holds

for moral judgment as “hard wired” in the brain. In other words, moral knowledge is “innate.” To explain moral knowledge is to explain how the brain exercises choice in making moral judgments. Under this explanation, making a moral judgment is a matter of actuating “the machinery to deliver moral verdicts based on unconscious and inaccessible principles.”¹¹ These principles, so the argument goes, are brought to bear on an ethical problem in a manner described as “unconscious.”¹²

The idea of unconscious rule following, grounded in the notion that moral knowledge is “encoded” or “embedded” in the brain, is a fundamental feature of the neurological explanation of human ethical judgment. As a form of explanation for human judgment, this approach suffers from some not insignificant flaws. To be clear, we are not contesting the empirical correctness of the view; we are saying that the view makes no sense, as such, as an explanation. Why does the idea of unconscious rule following make no sense? There are two reasons.

First, the idea of “tacit knowledge” has to be separated from that of “correct performance.”¹³ It is not enough to say that one’s brain “possesses” tacit knowledge because one performs correctly (i.e., in accordance with an ethical standard).¹⁴ Invoking tacit knowledge to

that ordinary individuals are intuitive lawyers, who possess tacit or unconscious knowledge of a rich variety of legal rules, concepts, and principles, along with a natural readiness to compute mental representations of human acts and omissions in legally cognizable terms”). Mikhail takes his cue from some remarks by John Rawls in *A Theory of Justice*.

¹¹ Marc D. Hauser, *Moral Minds* 42 (2006).

¹² See John Mikhail, John Mikhail, *Universal Moral Grammar: Theory, Evidence and the Future*, 11 *Trends in Cognitive Sci.* 143, 148 (2007) (arguing that moral knowledge is “tacit” and based on principles that are “unconsciously” applied).

¹³ See G.P. Baker & P.M.S. Hacker, *Wittgenstein: Understanding and Meaning* (Volume 1 of *An Analytical Commentary on the Philosophical Investigations*) 185 (2d ed., Revised by P.M.S. Hacker 2005) (“There must be identifiable conditions that will differentiate between possession of tacit knowledge and total ignorance, conditions distinct from correct performance.”).

¹⁴ We discuss what it means to “possess” knowledge in more detail in Section III.

explain behavior requires something more than the mere invocation to show exactly *how* tacit knowledge is doing the work claimed for it. If tacit knowledge is to be more than a question-begging explanation, there must be independent criteria for it. Lacking such criteria, the explanation is empty.¹⁵

Second, we question the intelligibility of the very idea of “unconscious” rule following. What can it possibly mean to say—as so many do—that the brain “follows rules unconsciously”? Rules are not causal mechanisms in the sense that they do not “act at a distance.”¹⁶ Rule following is something only human beings do and they do it not alone with their brains but in concert with others.

This last point can be detailed further. Consider that in many contexts in daily life where rules come into play, the following things seem to be implicated. We may (1) justify our behavior by reference to a rule; (2) consult a rule in deciding on a course of conduct; (3) correct our behavior and that of others by reference to a rule; and (4) interpret a rule when we fail to understand what it requires. Rule following occurs in a wide variety of contexts, each of which has its own unique features. These contexts are not “in the mind” (or “in the brain”) but in the world. They are referred to in the course of any explanation of what a subject thinks is required by a norm and what, on the subject’s view, that norm requires. When disputes break out about what norms require, appeal to what is in one’s head is question-begging, for the very presence of a different point of view on what a norm requires signals that the answer to the question cannot simply be summoned from the depths of the unconscious.¹⁷ Reason-giving in defense of a

¹⁵ One is here reminded of the explanation that opium puts people to sleep because it has “dormative powers.”

¹⁶ Baker and Hacker, at 186.

¹⁷ The same difficulty besets Moral Realism. For discussion, see Dennis Patterson, “Dworkin on the Semantics of Legal and Political Concepts,” 26 *Oxford J. Legal Studies* 545-557 (2006).

challenge about what a norm requires cannot be done from the realm of the unconscious. And once reasons are called for, the realm of the unconscious is left behind.

Moreover, there is a fundamental difference between following a rule and acting in accordance with a rule. Consider a simple example. In the entrance to a club in central London, the following sign appears on the wall: “Gentlemen are required to wear a jacket in the dining room.” Mr. Smith is a dapper man, who happens to be the guest of a club member. If Mr. Smith has his jacket on as he enters the dining room, we can safely say that he is “acting in conformity with the rule.” But is he “following” the rule? For that, more is required.

To actually “follow” the rule, Smith would have to be aware of it.¹⁸ If Smith had no knowledge of the rule prior to his entrance into the club, it is difficult to say how he was “following” the rule. How would he have conformed his conduct to the rule through a course of behavior (e.g., being told the dress code by his friend, the club member)? If Smith had his jacket on his arm and did not see the rule posted on the wall, he would not be acting in accordance with the rule and would, presumably, conform his conduct to the rule once he was apprised of it.

The point here is that there is an epistemic component to rule following: one has to be aware of the rules. Bringing one’s conduct in conformity with rules is an essential feature of “rule following.” Without this epistemic component, one is merely acting in accordance with what a rule requires. This, as we have argued, is not rule following in any meaningful sense.

¹⁸ See Maxwell Bennett and Peter Hacker, *The Conceptual Presuppositions of Cognitive Neuroscience*, in Maxwell Bennett, Daniel Dennett, Peter Hacker, and John Searle, *Neuroscience and Philosophy: Brain, Mind and Language* 151 (2007) (“But for something to constitute following a rule, the mere production of a regularity in accordance with a rule is not sufficient. A being can be said to be following a rule only in the context of a complex practice involving actual and potential activities of justifying, noticing mistakes, and correcting them by reference to the rule, criticizing deviations from the rule, and, if called upon, explaining an action in accordance with the rule and teaching others what counts as following the rule.”)

II. Interpretation

In the view of many neuroscientists and their enthusiasts, the brain does all manner of things. It describes, understands, computes, interprets, and makes decisions. In this section, we will focus our attention on one of these claims, to wit, that the brain achieves knowledge through a process of “interpretation.” While they are not alone in this regard, many scholars writing about neuroscience are enthusiastic in their belief that the brain grasps norms through an internal process of “interpretation.” Here is Oliver Goodenough singing the praises of Michael Gazzaniga’s “interpreter module” in the legal context:

[Gazzaniga] has postulated the existence of an interpreter module, whose workings are also in the word-based arena. A similar word-based reasoner could work with the word-based rules of law. In experiments on split-brain patients, whose central corpus callosum had been cut as a cure for severe epileptic problems, the interpreter supplied completely erroneous explanations for behavior originating in some nonword-based thinking module.¹⁹

Our problem with this account of mental life is that it fails to appreciate the fact that interpretation is a parasitic activity, one that is secondary to moral practice. Like Dworkin in legal theory,²⁰ neuroethicists want to make the case for interpretation as a fundamental feature of moral judgment. While we agree that interpretation is certainly an important element of both ethics and law, it is an activity that depends upon existing and widespread agreement in

¹⁹ Oliver R. Goodenough, Mapping Cortical Areas Associated with Legal reasoning and Moral Intuition, 41 *Jurimetrics J.* 429, 436 (2001).

²⁰ For discussion of this aspect of Dworkin’s jurisprudence and the dilemma posed for his theory of law, see Dennis Patterson, *Law and Truth* 71 – 98 (1996).

judgment. In short, interpretation cannot “get off the ground” without widespread agreement in judgment already being in place.

As Wittgenstein constantly reminds us, action is the ground of normativity and the nerve of the distinction between understanding and interpretation (“Im Anfang war die Tat”). The point of Wittgenstein’s example of the signpost in PI 85²¹ is that only action can provide the ground for correct and incorrect judgment.²² Without a practice of following it—a way of acting—the signpost by itself provides us no clue as to its proper use. In theory, there are as many potential ways of “following” the signpost as there are possible conventions for determining how it is to be used and what counts as following it. But once a convention for following signposts takes hold, a background of understanding evolves. It is against this background that the need for interpretation arises.²³ Interpretation is a reflective practice we engage in when understanding breaks down.

²¹ Wittgenstein, PI §85: “A rule stands there like a signpost. Does the signpost leave no doubt open about the way I have to go? Does it show which direction I am to take when I have passed it; whether along the road or footpath or crosscountry? But where is it said which way I am to follow it; whether in the direction of its finger or (e.g.) in the opposite one? -- And if there were, not a single signpost, but a chain of adjacent ones or of chalk marks on the ground – is there only one way of interpreting them? -- So I can say, the signpost does after all leave no room for doubt. Or rather: it sometimes leaves room for doubt and sometimes not. And now this is no longer a philosophical proposition, but an empirical one.”

²² All interpretation presupposes understanding. No one could interpret the following: *Nog drik legi xfo*m. The term first has to be translated or deciphered before interpretation takes place. Contra Quine, translation is not interpretation. W.V. Quine, *Ontological Relativity and Other Essays* (Columbia University Press, 1969) pp. 51-55. We interpret an utterance when we choose between *different* ways of understanding it. Legal interpretation is the activity of deciding which of several ways of understanding a rule given provision is the correct or preferable way of understanding. This is precisely the sort of activity Wittgenstein has in mind when he writes: “we ought to restrict the term ‘interpretation’ to the substitution of one expression of the rule for another.” Wittgenstein, PI at §201.

²³ See G.P. Baker and P.M.S. Hacker, *Wittgenstein: Understanding and Meaning* (Volume 2 of *An Analytical Commentary on the Philosophical Investigations*) (Blackwell, 1980) p. 667 (“[G]iving a correct explanation is a criterion of understanding, while the explanation given is a standard for the correct use of the expression explained. Correspondingly, using an expression in accordance with correct explanations of it is a criterion of understanding, while understanding an expression presupposes the ability to explain it.”).

Understanding is exhibited in action. For example, we show that we understand the request “Please shut the door” by closing the portal. The need for interpretation arises from a firmament of praxis.

As an account of correct and incorrect action in a practice (whether in ethics, law, arithmetic, or measurement), interpretation is a non-starter because interpretation draws our attention away from the techniques that make understanding possible. Correct and incorrect forms of action are immanent in practices. Correct forms of action cannot be imposed on a practice, by interpretation or otherwise. It is only when we master the techniques employed by participants in a practice that we can grasp the distinction between correct and incorrect action (e.g. in ethics or law – assertion).

We seek return to the quotidian dimension of praxis where the business of making claims about what morality requires are adjudicated through employment of intersubjectively employed standards of appraisal. Clarification of these standards and the rules for their use is the Wittgensteinian project expressed in jurisprudence. As Wittgenstein says, “It is not interpretation which builds the bridge between the sign and what is signified meant. Only practice does that.”²⁴

III. Knowledge

In the previous sections, we examined two issues that relate to particular kinds of knowledge: namely, what it means for a person to know how to follow a rule, and what it means for person to know (and to interpret) what is required of a norm. In this section, we turn to the concept of knowledge more generally. We first articulate a general conception of knowledge as

²⁴ Cited in G.P. Baker and P.M.S. Hacker, RGN, at 136.

a kind of ability, and we then apply this conception to legal examples that scholars have claimed neuroscience may inform.

The concept of knowledge has been a topic of intense philosophical interest for thousands of years, and understanding its contours is the main agenda for many epistemologists. Aside from theoretical issues epistemology, knowledge also relates to important ethical and practical issues. Ethical and legal judgments about whether to ascribe moral blame and/ or criminal responsibility to someone's actions often depend on what that person did or did not *know* when they acted, as well as what they were capable of knowing. Similarly, someone's knowledge of where they were and what they were doing on a particular occasion will virtually always be highly probative evidence of, for example, whether they are the perpetrator of a crime and ought to be held criminally liable. The promise that neuroscience might help us to determine conclusively what someone knows, or what they are/ were capable of knowing, is a seductive one.

In this section, we discuss some conceptual problems with claims of how neuroscience may inform questions of knowledge. To illustrate these problems, we will focus on four specific issues in law where significant consequences turn on whether a person is determined to know something. The legal issues fall into two main categories: a criminal defendant's mental states while committing allegedly criminal acts (concurrent knowledge), and whether a defendant has inculpatory knowledge of his previous criminal acts (knowledge of the past). These two categories each divide into two separate sub-issues with regard to questions of knowledge. With regard to a defendant's concurrent mental states, the two sub-issues are: (1) can neuroscience reveal whether the defendant was capable of having the requisite knowledge, and (2) if capable,

can it reveal whether the defendant in fact had the requisite knowledge?²⁵ With regard to the issue of knowledge of the past, the two sub-issues involve types of lie detection: (1) can neuroscience reveal whether a defendant has inculpatory knowledge in his brain, and (2) can it reveal whether a defendant is lying when he makes exculpatory assertions?

Before turning to these issues, we first outline a number of conceptual points regarding knowledge as a general matter. As with rule following and interpretation,²⁶ our fundamental methodological point is this: in order to assess the role that neuroscience may play in contributing to these issues, we must be clear about what knowledge is and what would count as someone having knowledge. More specifically, before we can determine whether someone knows something on a particular occasion, or is capable of knowing something more generally, we need some sense of the appropriate criteria for successful ascriptions of knowledge.²⁷

Ascriptions of knowledge generally take one of two forms: that someone knows *how* to do something (e.g., ride a bicycle, juggle, find one's way home, or recite the state capitals while juggling and riding a bicycle home) and that someone knows *that* a proposition is true ("that Springfield is the capital of Illinois," "that he lives on Sherwood Drive").²⁸ There is considerable overlap between these two types of knowledge ascriptions. Both *knowing-how* and

²⁵ As discussed below, these two questions are relevant to both the *mens rea* element of crimes and insanity defenses.

²⁶ See Sections I and II.

²⁷ In other words, we are not concerned with the empirical question of whether someone knows (or fits the criteria) on a particular occasion, but rather with the general criteria for ascriptions of knowledge.

²⁸ On the distinction between knowing-how and knowing-that, see Gilbert Ryle, *The Concept of Mind* 25-61 (1949). On the relationship between knowing-how and knowing-that, see Stephen Hetherington, "How to Know (that Knowledge-That is Knowledge-How)," in *Epistemology Futures* 71-94 (Hetherington ed., 2006).

knowing-that are manifested in successful behavior²⁹—in other words, in the ability to display the relevant knowledge. These manifestations—that is, expressions of knowledge—may take a variety of forms depending on the particular circumstances. You may manifest your knowledge of *how* to do something, for example, by doing it or by saying how it is to be done.³⁰ You may manifest your knowledge *that* something is true, for example, by asserting it, by answering questions about it correctly, by correcting others who are mistaken about it, or by acting appropriately based on that knowledge. It is also possible that you may do nothing at all with your knowledge (how or that). The primary point is that knowledge is a kind of cognitive achievement or success—it consists in a kind of power, ability, or potentiality possessed by the knowing agent.³¹

To be sure, this is not to suggest that knowledge just is the relevant behavior. As we noted, it is possible to have knowledge without expressing it. It is also possible to engage in the relevant behavior without in fact having knowledge. A lucky guess, for example, that something is true or how to do something is not knowledge.

Although knowledge is typically (but not always) manifested in behavior, one might object that certain types of syndromes or injuries pose a fundamental challenge to the conception of knowledge as an ability. Consider the tragic case of “locked-in syndrome,” in which victims, due to injury to their brain stems, remain fully conscious—with their memories and knowledge

²⁹ “To know” is an achievement or success verb; it refers to a kind of cognitive accomplishment.

³⁰ The latter might be the case for someone no longer physically capable of performing a task but who still knows how to do something (perform a dance or play a sport, for examples).

³¹ M.R. Bennett & P.M.S. Hacker, *A History of Cognitive Neuroscience* 96 (2008) (“To know something to be thus-and-so is ability-like, hence more akin to a power of potentiality than to a state or actuality.”)

intact—but are unable to move or talk.³² Plainly, they have knowledge—but they lack the ability to manifest their knowledge in the typical ways. Does this mean that knowledge is not, in fact, an ability, but rather is something else (a brain state)? We think not. First, those with locked-in syndrome can, quite remarkably, learn to communicate their knowledge through a series of complex eye movements.³³ These communications do manifest knowledge consistent with an ability conception of knowledge. And before a locked-in sufferer learns to communicate in this way—or in cases of “total locked-in syndrome” in which no movements of the eye or any other body parts are possible—he is still conscious of his knowledge and can recite it to himself and otherwise reflect on it. This too is an ability, and, indeed, it is part of the reason why we ascribe knowledge to patients in this situation. If such a patient were not conscious of their knowledge *in any way*, and could not manifest it in any way, on what basis would we ascribe knowledge to them? We would not. Thus rather than posing a challenge to the claim that the criteria for knowledge ascriptions includes an ability to manifest that knowledge in behavior, this example is consistent and reinforces that conception.

A second potentially challenging example is someone in a persistent vegetative state (PVS). This example raises two separate issues. First, someone in a PVS appears to have lost consciousness of their knowledge. Do they still possess knowledge? We think the best characterization of this situation is that they had knowledge and may later regain that knowledge when they resume consciousness. At best, they have the potential for knowledge.

Second, those in a PVS also sometimes do engage in behaviors that, under ordinary circumstances, might be consistent with manifestations of knowledge. For example, although

³² See Jean-Dominique Bauby, *The Diving Bell and the Butterfly* (1997). For a discussion of this syndrome and the questions it poses for neuroscience, see Noe, *supra* note 3, at 14-17.

³³ *Id.*

patients in this condition are thought to be unconscious, “it is not uncommon for [them] to respond to sounds, to sit up and move their eyes, to shout out, to grimace, to laugh, smile, or cry.”³⁴ Do they have knowledge? If they do not, but they have an ability to manifest responses to their environment, does this mean that knowledge is not an ability (to manifest such responses)? We think not. First, as noted above, one may engage in behavior that is consistent with knowing (how to do something or that something is so) without in fact possessing that knowledge (e.g., someone who answers a question correctly by guessing). The behavior, in other words, is not sufficient for knowledge.³⁵ Second, while knowledge implies an ability to do something, the reverse is not true: being able to do something does not imply knowledge.³⁶ The ability to do something may apply to many circumstances in which an ascription of knowledge is inappropriate. The ability of a metal to conduct electricity, for example, does not mean the metal *knows* how to conduct electricity. The ability of a thermometer to display the correct temperature does not mean the thermometer knows, for example, that it is currently 70 degrees. Knowledge, by contrast, requires a kind of “two-way ability,” in which agents may typically choose to or refrain from exercising at will.³⁷ With knowledge, as with rule following, it makes sense to say that an agent knows how to do something correctly, as well as what it means to do it incorrectly, to make a mistake, or to do it wrongly. To the extent someone in a PVS lacks consciousness, and thus this kind of two-way ability, their behavior may not qualify as knowledge.

³⁴ Noe, *supra* note 3, at 17.

³⁵ As noted above, the behavior is not necessary either—someone may have knowledge and choose not to express or manifest it in any way.

³⁶ Likewise, one may know how to do something but not be able to do so. On the distinction between “being able to” and “knowing how to” see Bennett & Hacker, *History*, *supra* note 31, at 97-99.

³⁷ On “two-way abilities” see *id* at 97-98.

Before turning to the legal examples, let us summarize the general discussion so far. We traced a general conception of knowledge as an ability, in which *knowledge-that* and *knowledge-how* may each be manifested or expressed in manifold ways based on the circumstances. We considered a couple of examples that may pose a challenge to this conception—lock-in syndrome and PVS—but, on the contrary, we found these examples consistent with our conception. We recognize, however, that some readers may disagree with our characterizations of these examples. Thus we are willing to concede that there may be penumbral cases (including these examples and others) in which reasonable people may disagree about how and when to ascribe knowledge. Nevertheless, we suggest that for typical, core cases of knowledge the best conception of knowledge is as a kind of ability. With this conception in place, we now turn to the legal examples.

A. Knowledge and Criminal Responsibility

A criminal defendant's state of mind is often relevant for determining whether a crime has been committed, which crime has been committed, and the appropriate level of punishment. Criminal laws typically define criminality in terms of prohibited acts (the *actus reus*) coupled with general or specific states of mind (the *mens rea*). In addition, defendants may raise additional state-of-mind issues as a defense, such as insanity or extreme emotional disturbance. With both *mens rea* and insanity defenses, the defendant's knowledge will often be a key determinate.

Consider first, the issue of *mens rea*. Knowledge comprises a common category of *mens rea* in the definition of various crimes. For example, The Model Penal Code specifies four levels

of mens rea: (1) Purposely (intent); (2) Knowingly; (3) Recklessly; and (4) Negligently.³⁸ A person acts “knowingly” with regard to an element of a criminal offense when “he is aware that his conduct is of that nature or that such circumstances exist.”³⁹ Also, if a crime is defined in terms of a result (e.g., “causing the death of another”), a person acts knowingly when “he is aware that it is practically certain that his conduct will cause such a result.”⁴⁰ Even when crimes are not explicitly defined in terms of knowledge—but rely on another level such as purpose/intent, recklessness, or negligence—the defendant’s knowledge is still relevant. One may act with criminal purpose or intent in many circumstances only when knows certain things. For example, one can commit theft intentionally only if one first knows that the property he is taking belongs to another. Similarly, a determination of whether one has acted “recklessly” or “negligently” may depend on the “circumstances known to him” at the time of his actions.⁴¹ Thus, virtually every *mens rea* determination—and thus whether a crime has been committed—will potentially involve a question of what the defendant knew (*knowledge-that*) at the time she acted.

Likewise, insanity defenses depend upon the defendant’s knowledge. Although the definitions vary somewhat across jurisdictions, common tests for insanity examine whether the defendant: failed to “know the nature and quality of the act he was doing”⁴²; failed to “know what he was doing was wrong”⁴³; failed to “appreciate the wrongfulness of his conduct”⁴⁴; or

³⁸ Model Penal Code § 2.02.

³⁹ Id. at §2.02(b)(i).

⁴⁰ Id. at §2.02(b)(ii).

⁴¹ Id. at §2.02(c)-(d).

⁴² See *Clark v. Arizona*, 548 U.S. 735, 747 (2006); *M’Nagten’s Case*, 8 Eng. Rep. 718 (1843).

⁴³ Id.

lacked the capacity to “conform his conduct to the requirements of the law”⁴⁵ (sometimes a lack of *knowledge-how*). Thus, as with *mens rea*, a criminal defendant’s concurrent knowledge will be crucial in determining insanity and thus criminal culpability.

How might neuroscientific evidence assist in answering disputed questions regarding a defendant’s concurrent knowledge? It has been suggested that it may do so in one of two different ways: either by revealing whether the defendant did or did not have the *capacity* to possess such knowledge⁴⁶ or by revealing more directly whether the defendant did or did not have such knowledge.⁴⁷ (Note that this neuroscientific evidence might serve the prosecution in establishing the relevant mental state or capacity or the defense in establishing the absence of the mental state or capacity.)

Given that knowledge is a kind of ability, the use of neuroscience to show a lack of capacity is much more plausible than using it to prove knowledge on a particular occasion. A properly working brain is a necessary condition for possessing knowledge. It is plausible that certain brain structures may be necessary in order to possess certain kinds of knowledge; thus, damages to these structures may make it the case that one lacks a capacity to possess some types of knowledge. This possibility raises a host of difficult empirical questions within the domain of neuroscience about brain plasticity and the like, but this empirical investigation—and the

⁴⁴ Model Penal Code

⁴⁵ *Id.*

⁴⁶ See Eyal Aharoni et al., “Can Neurological Evidence Help Courts Assess Criminal Responsibility? Lessons from Law and Neuroscience,” 1124 *Ann. N.Y. Acad. Sci.* 145-60 (2008).

⁴⁷ See Erin Ann O’Hara, “How Neuroscience might Advance the Law,” 359 *Phil. Trans. R. Soc. Lond.* 1677-1684 (2004).

possible subsequent juridical use of neuroscience for this purpose—is perfectly consistent with the conceptual issues with which we are concerned.

This is not so with claims about establishing *mens rea* directly. Putting aside the obvious practical limitation that we typically cannot scan the brains of defendants while they are committing crimes⁴⁸—what if we could? Suppose we had an fMRI scan of a defendant’s brain while committing allegedly criminal acts. Now, suppose we need to determine whether he (1) knew the suitcase he walked off with was not his but belonged to another; or (2) knew about the illegal drugs in his suitcase; or (3) knows “that it is practically certain that his conduct will cause” a particular result; or (4) knows any other particular proposition. Exactly where in his brain would we find this information? Because knowledge is an ability and not a state of the brain, the answer is: nowhere.

Given the practical limitations, the example may seem silly. But the conclusion and its implications are not. This becomes apparent when we turn to the issue of lie detection and claims about establishing a defendant’s knowledge of the past—an issue that presents real, immediate, critically important (and, as we will show, conceptually problematic) examples.

B. Knowledge and Lie Detection

One of the most prominent and most provocative proposed uses of neuroscientific evidence in law is to detect lies or deception. The effects of this use are potentially widespread because the neuroscientific evidence would be relevant to every witness and possible witness in every litigated and potential case. Use on criminal defendant’s—both to inculcate and to exculpate—however, is the example that receives the most attention. Proposed brain-based lie

⁴⁸ Perhaps the crime of perjury is an exception.

detection comes in two varieties: that attempting to discover if a defendant has knowledge of a crime “stored” or “housed” in his brain,⁴⁹ and that attempting to discover whether a defendant’s answers trigger areas of the brain correlated with deceptive behavior.⁵⁰ The first type attempts to discover the presence of knowledge directly; the second type attempts to infer knowledge on the basis of deception.

Consider, first, attempts to discover knowledge in the brains of criminal defendants. According to this approach, when presented with information that a defendant would know only if he were at a crime scene (e.g., how the victim was killed), certain brain activity will be triggered if he already knows or recognizes this information, whereas other brain activity will ensue if he is learning it for the first time.⁵¹ In characterizing the evidentiary value of this evidence, two proponents write: “the brain of the criminal is always there, recording events, in some ways like a video camera.”⁵² Examining the criminal’s brain, it is contended, will thus allow access to this “recording” and the underlying events of the crime.

This characterization depends on a confused conception of knowledge. To know something—knowledge that propositions about a crime are true, for example—is not located in the brain. As a conceptual matter, neural states of the brain do not fit the criteria for ascriptions of knowledge. Suppose, for example, a defendant has brain activity that is purported to be

⁴⁹ See Lawrence A. Farwell & Sharon S. Smith, “Using Brain MERMER Testing to Detect Knowledge Despite Efforts to Conceal,” 46 *J. Forensic Sciences* 135 (2001). A similar technique also provided the basis for a recent conviction in India, see See Anand Giridharadas, “India’s Novel Use of Brain Scans in Courts is Debated,” *NY Times*, Sept. 14, 2008.

⁵⁰ See, e.g., F. Andrew Kozel et al., “Detecting Deception Using Functional Magnetic Resonance Imaging,” 58 *Bio. Psychiatry* 605 (2005); D.D. Langleben, “Brain Activity during Simulated Deception: An Event-Related Functional Magnetic Resonance Study,” 15 *NeuroImage* 727 (2002).

⁵¹ See Farwell & Smith, *supra* note 49.

⁵² *Id.*

knowledge of a particular fact about a crime. But, suppose further, this defendant sincerely could not engage in any behavior that would count as a manifestation of knowledge. On what basis could one claim and prove that the defendant truly had knowledge of this fact? We suggest there is none; rather, evidence of the defendant's lack of the criteria for knowledge would override claims that depend on the neuroscientific evidence.⁵³

This confused conception of knowledge “stored” in the brain rests on two additional problematic assumptions. First, that memories (like knowledge) may be identified with particular neurological states of the brain. Second, an assumption that the *retention* of an ability implies the *storage* of that ability. Exposing these problematic assumptions further reveals what is wrong conceptually with claims made on behalf of this form of lie detection.

Memory is the retention of knowledge previously acquired or possessed by a person.⁵⁴ Like knowledge, memory, too, is an ability, and one that may be manifested in manifold ways. One may say what one remembers, think but not say what one remembers, act on what one remembers, and so on. Memories may be factual (e.g., remembering that Addis Ababa is the capital of Ethiopia), experiential (e.g., remembering seeing Ethiopia), or objectual (e.g., remembering the sight of one's room in Ethiopia).⁵⁵ In addition, memory may be declarative—that is, in propositional form, possibly true or false—or non-declarative (e.g., remembering how to ride a bicycle). In none of these varieties of memory is the criterion for whether one remembers that one has a particular neurological state. Rather, memory is the retention of

⁵³ The behavior provides “criterial” evidence of knowledge and the neuroscience provides “inductive” evidence of knowledge. If there is a discrepancy between them, the problem is with the inductive correlations; the brain activity, in other words, would not be well-correlated with knowledge.

⁵⁴ For further discussion of the conceptual contours of memory, see Bennett & Hacker, *History*, supra note 31, at 99-112.

⁵⁵ *Id.* at 100.

knowledge (an ability) and, like knowledge, the criteria include the various ways that this ability may be manifested in behavior. This is not to suggest that certain brain states and synaptic connections are not necessary for one to have the capacity to engage in this behavior, and understanding these conditions is an important avenue of neuroscientific research. But it is a mistaken leap to move from the fact that brain structures may be necessary for memory to the conclusion that memories are identical with such neurological states.⁵⁶

This leads to the second problematic assumption. Just because memory is knowledge retained, it does not follow that this retention is stored or housed in the brain (or elsewhere). One may possess an ability (knowledge, memory) without that ability being stored anywhere; indeed, it makes no sense to speak of storing an ability. The idea that the brain records and stores the events a person perceives (like a video camera), and that in memory one accesses this recording, is not a plausible explanation. First, it presupposes memory and cannot explain it because to make use of this neural record people would have to remember how to access and interpret it.⁵⁷ Second, even if such a neural record did exist, it would be unavailable (people typically cannot see their brains). To be sure, knowledge and memory casually depend on neurological states, and possessing and retaining knowledge will result in neurological changes, but it does not follow from this that the neurological phenomena “store” or “house” the knowledge or memories. Proposed lie detection that relies on such a problematic conception thus cannot succeed in showing what it purports to show.

⁵⁶ See Glannon, *supra* note 2, at 325 (“it is misleading to ask questions such as ‘Where in the brain is the memory of one’s past?’”).

⁵⁷ See Bennett & Hacker, *supra* note 31, at 107. Likewise, one can use a video recording to remember past events only if one remembers what the recording is a recording *of*.

A second type of brain-based lie detection is attempting to locate the neural correlates of deception.⁵⁸ Note that lies and deception are not necessarily the same thing: one can deceive without lying, and one can lie without deceiving (or even intending to deceive).⁵⁹ The scenarios that the neuroscientific studies are investigating typically include both. More specifically, test subjects are told to assert a false proposition in an attempt to deceive the scientists; meanwhile, the subjects' brains are scanned to examine whether there is increased activity in particular areas.⁶⁰ As with knowledge, these types of deceptive lies involve a complex ability engaged in by persons, not their brains.⁶¹ This behavior typically includes (1) recognizing the truth (or what is believed to be true); (2) choosing not to manifest their knowledge (or beliefs); (3) making assumptions about what the listener already believes and is likely to believe; and (4) uttering something false. This behavior provides the criteria for deceptive lies, not particular states of the brain. Thus, neuroscientific evidence might reveal that certain brain activity is inductively well-correlated with this behavior,⁶² or that damage to certain brain areas makes one incapable of engaging in this behavior, but it cannot establish conclusively that one's brain is engaged in lies or deception or that an intent to deceive or a lie is located in the brain. Neurological states do not fit the criteria for ascriptions of lies or deception. Thus, as with knowledge, if there are

⁵⁸ See supra note 50. See also <http://www.cephoscorp.com/> and <http://noliemri.com/>.

⁵⁹ For a discussion of the difference and illustrative examples, see Don Fallis, "What is Lying?" *CVI J. of Philosophy* 29 (2009).

⁶⁰ The fact that lies are associated with *increased* brain activity may be problematic. It may detect novice liars, but fail to catch experts, if—like athletes and musicians—expert liars appear to generate less of the relevant brain activity than novices. See John Milton et al., "The Mind of Expert Motor Performance is Cool and Focused," *NeuroImage* 804-13 (2007).

⁶¹ Although certain brain structures may be necessary to be capable of engaging in this behavior.

⁶² On the empirical limitations of current studies in establishing such correlations, see George T. Monteleone et al., "Detection of Deception Using Functional Magnetic Resonance Imaging: Better than Chance, But Well Below Perfection," *Social Neuroscience* (forthcoming).

discrepancies between the relevant behavior and neuroscientific evidence, the behavioral evidence will override the neuroscience.

Conclusion

We have examined the conceptual foundations of three activities: rule following, interpretation, and knowledge. Each type of activity is manifested in manifold types of behavior, which form the criteria for ascriptions of these activities. People, in various contexts, may engage in these types of behavior. Brains, by contrast, do not engage in these types of behavior. This is why we say that claims that the brain “follows rules,” “interprets,” or “knows” are lacking in sense. We are not saying that such claims are false—brains do not break the rules, misinterpret, or are ignorant. Rather, given that brains do not engage in the types of behavior that form the criteria for these concepts, we do not know what it *means* to make such claims about brains or brain regions. The normal senses in which we use “follow a rule,” “interpret,” and “know,” seem not to apply. Through these examples, we have attempted to show that claims regarding how neuroscience may inform issues in ethics or law must make sense before they can be true.