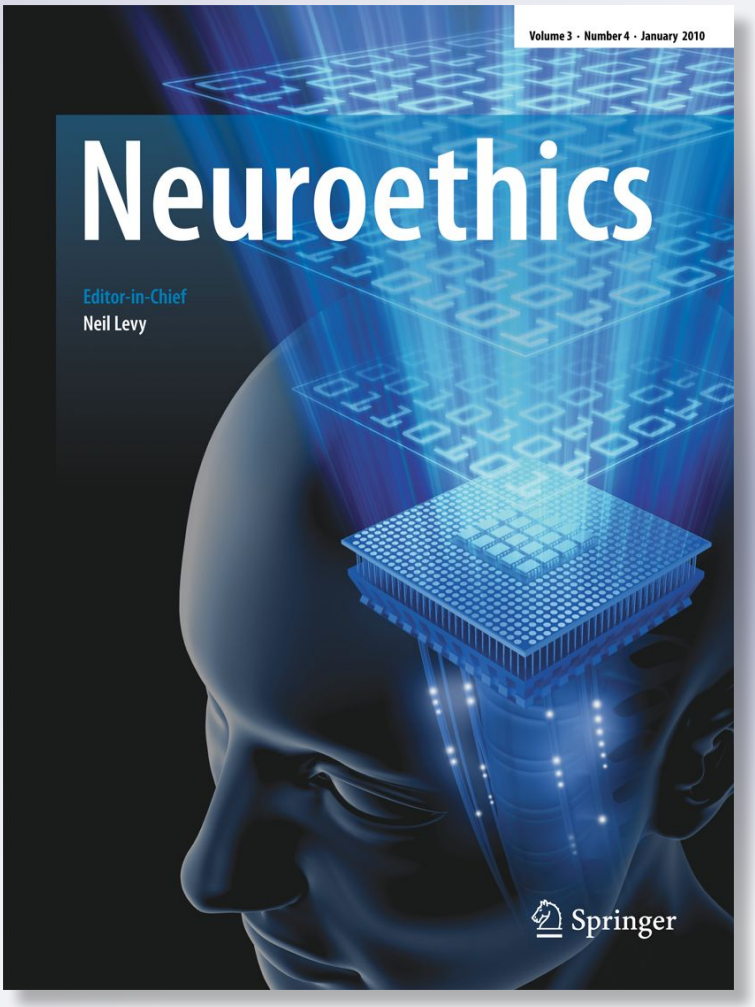
Neural Lie Detection, Criterial Change, and OrdinaryLanguage

Thomas Nadelhoffer

Neuroethics

ISSN 1874-5490
Volume 4
Number 3

Neuroethics (2011) 4:205-213
DOI 10.1007/s12152-010-9080-6



Springer

Neuroethics (2011) 4:205–213
DOI 10.1007/s12152-010-9080-6

ORIGINAL PAPER

Neural Lie Detection, Criterial Change, and Ordinary Language

Thomas Nadelhoffer

Received: 17 March 2010 / Accepted: 17 April 2010 / Published online: 29 June 2010

Abstract Michael Pardo and Dennis Patterson have recently put forward several provocative and stimulating criticisms that strike at the heart of much work that has been done at the crossroads of neuroscience and the law. My goal in this essay is to argue that their criticisms of the nascent but growing field of neurolaw are ultimately based on questionable assumptions concerning the nature of the ever evolving relationship between scientific discovery and ordinary language. For while the marriage between ordinary language and scientific discovery is admittedly not always a happy one, it is an awkward union that nevertheless seems to work itself out with the passage of time. In the following pages, I will try to show that Pardo and Patterson's primary argumentative strategy ultimately depends on basic assumptions concerning the fixity of language that we should reject.

Keywords Neurolaw · Ordinary language · Scientific discovery · Lie detection

This material is based upon work supported by the John D. and Catherine T. MacArthur Foundation, the Law and Neuroscience Project, and The Regents of the University of California. So, I greatly appreciate their generous support. I would also like to thanks Dickinson College for providing me with a leave of absence so that I could be a Law and Neuroscience Project post-doctoral fellow.

T. Nadelhoffer (✉)
Dickinson College,
Carlisle, PA, USA
e-mail: tnadelhoffer@gmail.com

> *"But is there then no objective truth? Isn't it true, or false, that someone has been on the moon?" If we are thinking within our system, then it is certain that no one has ever been on the moon. Not merely is nothing of the sort ever seriously reported to us by reasonable people, but our whole system of physics forbids us to believe it. For this demands answers to the questions "How did he overcome the force of gravity?" "How could he live without an atmosphere?" and a thousand others which could not be answered.*
> –Ludwig Wittgenstein, *On Certainty* ([1], §108)

Introduction

Michael Pardo and Dennis Patterson have recently put forward several provocative and stimulating criticisms that strike at the heart of much work that has been done at the cross-roads of neuroscience and the law.[1] Taking the philosophy of Ludwig Wittgenstein as their starting point and building upon recent work by Max Bennett, Peter Hacker, Alva Noe and others,[2] Pardo and Patterson set out to challenge the purport-

[1] See, also, Pardo and Patterson [2].
[2] See, e.g., Hacker [3]; Bennett and Hacker [4]; Murphy and Brown [5]; Morse [6, 7]; Noe [8].

edly conceptually confused framework that they claim underlies many of the recent attempts to apply data from cognitive neuroscience to particular legal issues ranging from free will to lie detection. My goal in this essay is to argue that their criticisms of the nascent but growing field of neurolaw[3] are ultimately based on questionable assumptions concerning the nature of the ever evolving relationship between scientific discovery and ordinary language. As the quote from Wittgenstein above illustrates, yesterday's impossibilities have a way of becoming tomorrow's platitudes. For while the marriage between ordinary language and scientific discovery is admittedly not always a happy one, it is an awkward union that nevertheless seems to work itself out with the passage of time. In the following pages, I will try to show that Pardo and Patterson's primary argumentative strategy ultimately depends on basic assumptions concerning the fixity of language that we should reject.[4] If I am right, then Pardo and Patterson still have some additional argumentative spadework to do before we should accept their skeptical conclusions concerning both neural lie detection and neurolaw more generally.

Pardo and Patterson and the Primacy of Behavioral Criteria

Before we examine the intended targets of Pardo and Patterson's criticisms, it will be helpful to identify the views they explicitly claim *not* to be challenging or questioning. First, they are not denying that particular neurological states, processes, and events, "may be a necessary condition for various mental activities." However, on their view, even if brain states are *necessary* for specific patterns of behavior that we take to be essential for personhood or agency, these brain states are not in themselves *sufficient* for personhood or agency. For present purposes, I am going to call the view that Pardo and Patterson are rejecting here the reductive sufficiency thesis—which is a view that we will revisit in §3. Second, Pardo and Patterson's main target is not *neuroscience itself*—viewed broadly as the collective attempt to shed light on "the brain, mind, and human behavior, and the relationship between them"—but rather the sorts of things that some researchers are tempted to say when fleshing out the *implications of neuroscience*.

Much like Wittgenstein before them, Pardo and Patterson set out to expose the assumptions which they believe tempt us to make claims that are conceptually confused. On their view, by focusing our attention on the kinds of *behavioral* criteria that we ordinarily rely on when talking about persons, minds, knowledge, and the like, we will see that the kinds of statements made by proponents of the reductive sufficiency thesis—e.g., "you are your brain"[5]—are not *false* but rather *lacking sense altogether*. For present purposes, I am going to call this conventionalist element of Pardo and Patterson's view the primacy of behavioral criteria. Because this is an assumption that plays a pivotal role in their overall argument, we should pause briefly to consider an illustrative example.

Consider, for instance, smiling. If I want to say something meaningful about smiling, and if the criteria for the correct application of the concept are behavioral, then it will only make sense to talk about smiling in contexts within which these behavioral criteria could be met. So, while it makes sense to say that 'Paige is smiling'—i.e., we know what it means to see whether the criteria for smiling have been satisfied in her case—it does not make sense to say that "Paige's tooth is smiling."[6] After all, there are no criteria for determining

[3] It is worth pointing out that the term "neurolaw" refers to a growing interdisciplinary *field of inquiry* that explores the relationship between neuroscience and the law. As such, neurolaw is neither an intellectual movement nor is it tied to a certain ideological point of view. Just as some researchers push a revolutionary agenda whereby neuroscience ought to be used to overthrow traditional legal categories (see, e.g., Greene and Cohen [9]), other researchers adopt a much more conservative approach to the relationship between law and neuroscience (see, e.g., Morse [6]). Given this very wide spectrum of views, it is a mistake to identify the overall field of neurolaw with particular positions within the field.

[4] I think the argument could even be made that we have Wittgensteinian grounds for resisting the conventionalism of Pardo and Patterson, but that will not be one of my direct goals in this commentary.

[5] The claim by Joshua Greene and Jonathan Cohen that "you are your brain" [9, p. 1779] is identified by Pardo and Patterson as one of their primary targets. We will unpack Greene and Cohen's views concerning the relationship between the self and the brain in more detail in §3.

[6] We may sometimes talk about a "happy tooth"—e.g., if we just had an aching tooth repaired—but here we are using "happy" in a very loose way. In a similar way, we may talk about Paige's "smiling on the inside" as well—but here again, we would be using "smiling" very loosely. Moreover, the criteria we would rely on in trying to ascertain whether Paige really is "smiling on the inside" would themselves be behavioral criteria.

the truth of this latter statement. Teeth don't smile. Only people smile. To suggest otherwise is to purportedly succumb to what Bennett and Hacker have dubbed the "mereological fallacy"—i.e., the tendency to "ascribe properties to a part of the animal…that make sense only when ascribed to the animal as a whole" [4, p.73].[7] In describing the fallacy in more detail, Bennett and Hacker make the following pointed remarks:

> It is not semantic inertia that motivates our claim that neuroscientists are involved in various forms of conceptual incoherence. It is, rather, the acknowledgement of the requirements of the logic of psychological expressions. Psychological predicates are predicable only of a whole animal, not of its parts. No conventions have been laid down to determine what is to be meant by the ascription of such predicates to a part of an animal, in particular to its brain. So the application of such predicates to the brain… transgresses the bounds of sense. The resultant assertions are not false…Rather, the sentences in question lack sense. This does not mean that they are silly or stupid. It means that no sense has been assigned to such forms of words, and that, accordingly, they say nothing at all, even though they look as they do [4, p. 78].

Ultimately, Pardo and Patterson follow Bennett and Hacker in appealing to something roughly along the lines of what I am calling the primacy of behavioral criteria in an effort to establish that statements such as "Paige's brain knows, decides, intends, etc." are just as senseless as "Paige's tooth is smiling." Here again, brains don't know, decide, or intend. Only people and other whole animals do. To think otherwise is to be mired in confusion.

The general conceptual distinction that Pardo and Patterson encourage us to draw between what *people* can do and what *brains* can do has potentially wide ranging implications. However, for present purposes, we are going to limit our attention to what Pardo and Patterson have to say more specifically about recent attempts to use neuroscience for the purposes of lie detection. Once we have briefly examined their views on this front in §2, we will move on in §3 to discuss a few potential problems for Pardo and Patterson's general approach.

Knowledge and Lie Detection

According to the Wittgensteinian analysis of knowledge adopted by Pardo and Patterson,[8] both *knowing-how* and *knowing-that* are "manifested in successful behavior—in other words, in the ability to display the relevant knowledge."[9] So, on their view, while particular neurological states might be necessary for knowing-how or knowing-that, knowledge itself cannot simply be *reduced* to these neurological states. Instead, meaningful ascriptions of knowledge are intimately tied to behavioral criteria that neither whole brains nor individual neurological states could possibly satisfy. Given that Pardo and Patterson view knowledge as an *ability* that can only be expressed behaviorally by the whole animal, it is unsurprising that they adopt a skeptical stance towards the very possibility of neuroscience-based lie detection—which will be the topic of the remainder of this section.[10]

The first thing that needs to be pointed out is that Pardo and Patterson focus on what they take to be the two main varieties of neural lie detection that have

[7] Bennett and Hacker call the principle that is purportedly being violated "the mereological principle." As they say, "We have bluntly asserted the mereological principle in neuroscience, insisting that it is a logical principle, and therefore not amenable to empirical, experimental, confirmation or disconfirmation. It is indeed a convention, but one that determines what does and does not make sense. Its application—for example, to psychological concepts—could, in principle be changed by stipulation, but not without changing a great deal else, thereby altogether changing the meanings of our words and the structure of the multitude of familiar concepts. For the principle that psychological predicates apply to the animal as a whole and cannot be applied to its parts is held in place by a ramifying network of conceptual connections." [4, p. 81]. This is an issue that has recently been discussed at length by Noe [8] as well.

[8] See, e.g., Wittgenstein [1, 10].

[9] Pardo and Patterson are careful to point out that they are not suggesting that knowledge "just is the relevant behavior"—since it is clearly possible both to (a) "have knowledge without expressing it," and to (b) "engage in the relevant behavior without in fact having knowledge." But if satisfying the behavioral criteria is neither necessary nor sufficient for knowledge, it is unclear why these criteria ought to be exclusively used to delineate what can meaningfully be said about knowledge.

[10] Neuroscience-based lie detection is also sometimes called brain-based lie detection. In this paper, I am simply going to use "neural lie detection" for short.

been developed for the purposes of the law.[11] The first technique is a kind of neural familiarity test that uses electroencephalography (EEG) or functional magnetic resonance imaging (fMRI) to try to determine whether or not people are familiar with pertinent facts about the crime. The second technique uses fMRI to determine whether regions of the brain that have been shown in the laboratory to be associated with intentional acts of deception are activated when a defendant or witness is interviewed while being scanned. This latter approach is driven by the assumption that once we understand the neural correlates of honesty and deception, we will then be able to use neuro-imaging to determine whether legal actors are being deceptive.[12]

At the end of the day, Pardo and Patterson dismiss both forms of neural lie detection as confused rather than merely impractical or unreliable—i.e., their objection is *conceptual* rather than *empirical*. Indeed, Pardo and Patterson can entirely side-step the methodological, moral, and legal objections that ordinarily crop up in discussions about neural lie detection.[13] On their view, it's not that using neuroscience to detect lies is difficult or that it is likely to be more prejudicial than probative in the courtroom. Rather, the claim is that the entire project is confused from the start. Neither whole brains nor neural processes lie. Only people lie. As such, we can no more use neuroscience to detect lying in the brain than we can use dentistry to determine whether a patient's tooth is smiling. If Pardo and Patterson are right about this, neural lie detection truly is a non-starter. However, as we are about to see, it is unclear that we should accept without further argumentation some of the assumptions that are needed to get their view off the ground.

Ordinary Language and Scientific Discovery

As we have just seen, Pardo and Patterson have developed a general methodological framework for examining the kinds of statements that people sometimes make when talking about the relationship between neuroscience and the law. More specifically, they argue that ascriptions of mental abilities such as knowing, remembering, and lying are governed by behavioral criteria that neither whole brains nor neurological states could possibly satisfy or instantiate. In this section, I am going to try to show that Pardo and Patterson have not done enough to adequately motivate this first step in their argument. And if it turns out that we have reason to reject the view that behavioral criteria are the ultimate arbiters of sense when it comes to ascriptions of knowledge, memory, and deception, then the rest of Pardo and Patterson's argument will be on shakier ground. But let's not get ahead of ourselves.

Our first order of business at this point is to make sure we have a basic understanding of the Wittgensteinian notion of criteria that plays such an important role in Pardo and Patterson's overall argument. On their view, criteria "establish the norms for ascriptions of these concepts." As such, if x is a criterion for some event or state of affairs y, then when x obtains, one has defeasible grounds for concluding that y obtains. As Wittgenstein suggested in *The Blue Book*, defining criteria "give our words their common meaning" [37]. This is an aspect of Wittgenstein's view which Carol Caraway helpfully captures in the following way:

> Criterial rules of language fit the following schema: given certain general facts of nature, it is a criterial rule of language that in the appropriate particular circumstances, a certain type of behavior (B) is a criterion of P and in normal particular circumstances, someone's exhibiting an adequate array of criteria for P shows us (or justifies our assertion) that he is in P [38, p. 162].

[11] There are actually at least five distinct methods that are presently being developed that use neuroscience in one form or another for the purposes of lie detection. See Greely [11, p.48] for a discussion of these methods as well as their respective shortcomings. To date, there have been a limited number of peer reviewed studies on neural lie detection. Pardo and Patterson mention Kozol et al. [12] and Langleben et al. [13]. See, also, Davatzikos et al. [14]; Ganis et al. [15]; Langleben et al. [16]; Lee et al. [17]; Mohamed et al. [18]; Nunez et al. [19]; Spence et al. [20].

[12] The most recent study on neural lie detection—and arguably the most promising—is found in Greene and Paxton [21]. Their experimental design addresses several of the most prominent shortcomings of previous attempts to use fMRI for purposes of detecting honesty and deception.

[13] See, e.g., Farah and Wolpe [22]; Garland and Glimcher [23]; Greely [11]; Greely and Illes [24]; Kanwisher [25]; Kittay [26]; Langleben [27]; Moreno [28]; Morse [7]; Phelps [29]; Rakoff [30]; Schauer [31]; Sinnott-Armstrong et al. [32]; Spence [33]. For discussions of neural lie detection in the popular press, see Henig [34]; Narayan [35]; Silberman [36].

Hence, to say that writhing and moaning are behavioral criteria for the concept of pain is to say that if you see someone writhing and moaning, you have defeasible grounds for ascribing pain to that person. In this sense, criteria have normative force. However, we must be careful in this context to distinguish these kind of behavioral criteria from what Wittgenstein called symptoms—i.e., non-criterial states, processes, or events that we have learned purely through induction happen to correlate with certain criteria. Unlike criteria, which in some important sense fix the meanings of our terms, symptoms merely provide us with evidence concerning whether the criteria themselves have been met.

The primary example Wittgenstein uses when he introduces the distinction between criteria and symptoms is the case of angina—which we would now simply call "influenza" or "the flu." His extended remarks concerning the distinction are as follows:

> Let us introduce two antithetical terms in order to avoid certain elementary confusions: To the question "How do you know that so-and-so is the case?", we sometimes answer by giving '*criteria*' and sometimes by giving '*symptoms*'. If medical science calls angina an inflammation caused by a particular bacillus, and we ask in a particular case "why do you say this man has got angina?" then the answer "I have found the bacillus so-and-so in his blood" gives us the criterion, or what we may call the defining criterion of angina. If on the other hand the answer was, "His throat is inflamed", this might give us a symptom of angina. I call "symptom" a phenomenon of which experience has taught us that it coincided, in some way or other, with the phenomenon which is our defining criterion. Then to say "A man has angina if this bacillus is found in him" is a tautology or it is a loose way of stating the definition of "angina". But to say, "A man has angina whenever he has an inflamed throat" is to make a hypothesis.
>
> In practice, if you were asked which phenomenon is the defining criterion and which is a symptom, you would in most cases be unable to answer this question except by making an arbitrary decision ad hoc. It may be practical to define a word by taking one phenomenon as: the defining criterion, but we shall easily be persuaded to define the word by means of what, according to our first use, was a symptom. Doctors will use names of diseases without ever deciding which phenomena are to be taken as criteria and which as symptoms; and this need not be a deplorable lack of clarity. For remember that in general we don't use language according to strict rules-it hasn't been taught us by means of strict rules, either [37, p. 25].

The case of influenza is an especially interesting one for our present purposes. After all, as scientists started to understand the nature of the flu, what were once treated as regal criterial rules of language were subsequently demoted to the status of mere symptom—e.g., sore throat, fever, shivering, coughing, and the like. That this kind of criterial change is possible at all should give us pause when it comes to the conventionalist framework of Pardo and Patterson's argument. As James Klagge has pointed out in a similar context, "If change in the criteria of concepts…is possible, and Wittgenstein admits that it is, then what is to prevent neuroscience from discovering enough about brain states that we should eventually see it as natural to treat brain states as criteria for mental states and treat behavior as symptoms" [39, p. 323].

Fortunately, we need not concern ourselves here with the difficult exegetical project of pinning down precisely what Wittgenstein had in mind on this front.[14] There is a large literature on both the nature and the implications of Wittgenstein's notion of criteria—especially when it comes to the relationship between criterial change and scientific progress.[15] Whereas some portray Wittgenstein as a misguided and perhaps even inconsistent conventionalist (see, e.g., [43, 45, 48]), others think Wittgenstein has the tools to both accommodate and explain criterial change in the face of scientific progress (see, e.g., [38, 40]). Exploring the sprawling secondary litera-

[14] I agree with Koethe [40] that "For all the use Wittgenstein makes of the notion of criteria, he offers very little in the way of an explanation of it" (p. 603).

[15] See, e.g., Albritton [41]; Caraway [38, 42]; Chihara and Fodor [43]; Garver [44]; Hollinger [45]; Kenny [46]; Koethe [40]; Malcolm [47]; Putnam [48]; Scriven [49]. For a review of the early literature on the Wittgensteinian notion of criteria, see Lycan [50].

ture on Wittgensteinian criteria would take us too far afield. For present purposes, we should focus instead on whether Pardo and Patterson have adequately motivated their own version of criterial conventionalism. In the following pages, I will argue that they have not.

For starters, as we have just seen, it is unclear that the criteria of ordinary language are as rigid as Pardo and Patterson would need them to be for their argument to work effectively. Consider, for instance, what they say specifically about knowledge and memories being stored in the brain. On their view, brains *do not* and *could not* satisfy the behavioral criteria for meaningful ascriptions of knowledge, memory, etc. As such, it doesn't make any sense to talk about what the brain knows or remembers. Moreover, Pardo and Patterson claim that it is similarly mistaken to think that knowledge and memories are stored in the brain. On the one hand, they deny that it is possible to identify memories with particular neurological states given that the behavioral criteria don't apply at the neural level. On the other hand, they claim that since (a) both knowledge and memory are abilities, and (b) abilities cannot be stored anywhere, then (c) knowledge and memories could not possibly be stored in the brain. To think otherwise is to purportedly once again transcend the bounds of ordinary language.

But what is the supporting evidence for this claim about the criteria for ordinary ascriptions of knowledge? Indeed, I suspect that if you were to ask people on the street today where memory and knowledge are stored, the overwhelming majority would say "in the brain." Of course, whether I am right about this is an empirical question that calls for controlled and systematic investigation.[16] But for now, let's assume for the sake of argument that people do in fact find it entirely sensible to say that knowledge and memories are stored in the brain. How would that affect Pardo and Patterson's conventionalist criticisms of neural lie detection? If ordinary language is already trending in the reductive direction in light of recent developments in neuroscience, then this puts serious pressure on some of Pardo and Patterson's key claims.

After all, to the extent that ordinary language has the fluid capacity to accommodate scientific discovery, why we should follow Pardo and Patterson in thinking that the final arbiters of sense are always limited to the traditional behavioral criteria which happen to be presently in place? It would be like telling Alexander von Humboldt that "water is H_2O" lacks sense because molecules are not included in the criteria we ordinarily rely on when talking about water.[17] Just because the traditional criteria did not include events at the molecular level, it doesn't mean that we can't adopt new criteria for talking about water in light of developments in physics and chemistry. Similarly, just because the criteria we traditionally relied on when talking about mental activities such as knowing, deciding, intending, and lying were *behavioral*, it doesn't follow that *neural* criteria could not possibly be adopted *in the future* in light of developments in neuroscience. To assume otherwise is to overlook the relative fluidity of ordinary language.[18] Moreover, this general worry about whether Pardo and Patterson's conventionalism can adequately accommodate the real world relationship between ordinary language and science will be even more pressing for Pardo and Patterson if it turns out that ordinary language is already trending towards the very kinds of reductive ascriptions about the relationship between the mind and the brain that they dismiss as lacking sense.

For now, however, I want to set this worry aside and turn our attention instead once again to the reductive sufficiency thesis—a view that we are told

[16] Philosophers and psychologists who work in the nascent field of experimental philosophy often probe precisely these kind of folk intuitions with an eye towards shedding light on first-order philosophical problems. For general introductions to experimental philosophy, see Knobe [51]; Knobe and Nichols [52]; and Nadelhoffer and Nahmias [53].

[17] The general issue I am highlighting here was the motivating issue behind the influential debate between Norman Malcolm and Hilary Putnam concerning the relationship between criteria, ordinary language, and scientific discover. See, e.g., Hollinger [45]; Kenny [46]; Malcolm [47]; Putnam [48]. But since Pardo and Patterson did not frame their criticisms of neurolaw in terms of this salient earlier debate, I will set aside the details for now.

[18] Obviously, language cannot be too fluid. There need to be some rules that stand firm so that others can change. The issue we are talking about here, however, is not about the limits of language's fluidity. Instead, we are merely interested in whether the criteria of ordinary language are capable of change, expansion, or even fundamental revision.

is exemplified by Greene and Cohen's claim that "you are your brain."[19] Unsurprisingly, Pardo and Patterson reject this kind of reductive claim on the familiar grounds that it is lacking sense. After all, *you* are a person who thinks, knows, believes, and occasionally deceives. But your *brain* purportedly can't do any of these things. Hence, you cannot possibly be just your brain. As such, Pardo and Patterson conclude that Greene and Cohen have confusedly attributed mental properties to the brain that only a whole person could satisfy. But I am unsure that this is the most helpful or charitable way to interpret reductive claims such as "you are your brain."[20]

On my reading of Greene and Cohen, they are simply trying to point out that once we dispense with dualism—i.e., the view that the mental and the physical are ontologically distinct—we are left with the view that "every mental state and every difference in behavioral tendency is a function of some kind of difference in the brain" [9, p. 1779]. So, when Greene and Cohen say "you are your brain," they are not claiming that you just are your brain, *full-stop*. Instead, they are suggesting that the gathering data from neuroscience are not consistent with the traditional libertarian picture of the mind and agency whereby the conscious self somehow sits above the causal fray and makes decisions and initiates actions ex nihilo. It's not that there is absolutely nothing more to who you are than your brain. Instead, Greene and Cohen are merely pointing out that "what neuroscience does, and will continue to do at an accelerated rate, is elucidate the 'when', 'where' and 'how' of the mechanical processes that cause behavior" [9, p. 1781]. Moreover, they suggest that as neuroscience continues to make progress on this front, it will become increasingly clear that our mental lives are merely the byproduct of a "mass of neuronal instrumentation." As such, Greene and Cohen predict that folk intuitions concerning free will, agency, and responsibility will shift away from the dualism and retributivism that have historically held sway.

In some important sense, what Greene and Cohen say concerning the relationship between neuroscience and folk intuitions about agency and responsibility dovetails nicely with the worry I raised earlier about Pardo and Patterson's reliance on the primacy of behavioral criteria. After all, if Greene and Cohen are correct in assuming that developments in neuroscience could change how we envision the fundamental relationship between the brain and the main, then Pardo and Patterson's objections lose much of their force. In order to dismiss statements like "Paige's brain is lying" as lacking in sense because they fail to satisfy the behavioral criteria of ordinary language, Pardo and Patterson need these criteria to be fairly rigid. But then they owe us an argument that explains how it is possible for language to change in light of scientific progress even though the criteria that govern how we can meaningfully talk about the world are fixed. At this point, I minimally think that there are a sufficient number of historical counter-examples to the primacy of behavioral criteria to call Pardo and Patterson's use of it into question.[21]

Before closing, however, I want to briefly make one more observation about the specific way that Pardo and Patterson have framed the debate about neural lie detection. Keep in mind that on their view, only people can behave deceptively. So, while neurological processes may be necessary for lying and deception, the lies themselves are not located in the brain. Indeed, they are not located anywhere. According to Pardo and Patterson, "deceptive lies involve a complex ability engaged in by persons, not their brains." Because lies cannot possibly be "in the brain," it is a conceptually confused exercise in futility to look for them there. But is this what the researchers who are working on neural lie detection are really trying to accomplish—i.e., are researchers really trying to find lies in the brain?

[19] The complete quote is as follows: "It is not as if there is a you, the composer, and then your brain, the orchestra. You are your brain, and your brain is the composer and the orchestra all rolled together. There is no little man, no 'homunculus', in the brain that is the real you behind the mass of neuronal instrumentation" [9, p. 1779].

[20] The main issue I am interested in here is not merely exegetical. As such, my concern is not so much with what Greene and Cohen "really meant" but rather whether Pardo and Patterson have appropriately understood the real thrust of the reductive views they reject.

[21] It is worth pointing out that it is true that ordinary language historically relied heavily—if not exclusively—on behavioral criteria. As such, it is unsurprising that so many concepts have the sorts of criteria highlighted by Pardo and Patterson. However, the issue is not the ubiquity of behavioral criteria when it comes to ordinary language. Rather, the issue is whether we ought to use the present behavioral criteria as the sole and definitive normative guideline for distinguishing sense from non-sense.

Consider, for instance, the polygraph exam—an admittedly unreliable tool for the purposes of lie detection, but a good example for present purposes. How does it work? Does a polygraph exam actually reveal the lie itself? Of course not. A polygraph exam merely gives us information about an individual's physiological reactions to certain statements and questions—reactions that we take to be correlated with acts of intentional deception. When polygraph exams happen to work—which isn't as often as one would like—they provide us with *indirect* evidence of deceptive lying. But the same could be said about *all of the methods* of lie detection that have been developed thus far. Whether researchers are using (a) EEG to detect P300 wave activity, (b) periorbital thermography to detect increased temperature around the eyes, (c) near-infrared laser spectroscopy to create and record "scatter" patterns, or (d) fMRI to detect BOLD signals in certain regions of the brain, they are searching for the physiological *markers* or *signatures* of lying and deception. In each case, researchers are not interested in finding the lie itself in the brain. They are merely interested in gathering data that might be probative for the purposes of ascertaining whether someone is telling the truth.

As such, lie detection is technically a bit of a misnomer. No one really thinks that lying just is having an increased heart rate any more than people think that lying just is activation in a specific region of the brain. Lying is a very complex social behavior that gets played out under what are often very complicated circumstances. So, the claim is not that token lies are somehow hidden in the brain. Rather, the claim is that (a) lying is a complex behavior that requires certain cognitive processes (e.g., imagination and intention), and (b) lying is a complex behavior that tends to produce certain physiological side effects (e.g., nervousness and anxiety). The goal of neural lie detection is therefore simply to discover the neural correlates and physiological side effects of intentional acts of deception so that we might develop more reliable techniques for identifying people who are not telling us the truth.

I, for one, think that the gathering data concerning neural lie detection suggests that it is not only logically possible but also empirically feasible. As such, I think Pardo and Patterson have more argumentative spadework to do before we should follow their skeptical lead when it comes to the future of neural lie detection and neurolaw more generally. In the meantime, I think that ordinary language has already begun shifting towards precisely the kind of mechanical picture of the mind and the brain that Pardo and Patterson reject—a trend that I suspect will only continue as neuroscience progresses. Whether this is ultimately a healthy trend is a story for another day. For now, the important point is that the sometimes rocky relationship between ordinary language and science is an important topic that is ripe for future conceptual and empirical work at the cross-roads of neuroscience, philosophy, and the law.

References

1. Wittgenstein, L. 1969. *On certainty*. Oxford: Blackwell. Translated by G.E.M. Anscombe and Georg H. von Wright.
2. Pardo, M., and D. Patterson. 2010. Philosophical foundations of law and neuroscience. *Illinois Law Review*, forthcoming.
3. Hacker, P.M.S. 2007. The relevance of Wittgenstein's philosophy of psychology to the psychological sciences. *Proceedings of the Leipzig Conference on Wittgenstein and Science*.
4. Bennett, M.R., and P.M.S. Hacker. 2003. *Philosophical foundations of neuroscience*. Oxford: Blackwell.
5. Murphy, N., and W.S. Brown. 2007. *Did my neurons make me do it?* New York: Oxford University Press.
6. Morse, S. 2007. The non-problem of free will in forensic psychiatry and psychology. *Behavioral Sciences & the Law* 25: 203–220.
7. Morse, S. 2009. Actions speak louder than images. In *Using imaging to identify deceit: Scientific and ethical questions*, ed. Emilio Bizzi et al., 23–34. Cambridge: American Academy of ARts and Sciences.
8. Noe, A. 2009. *Out of our heads*. New York: Hill and Wang.
9. Greene, J.D. and J.D. Cohen 2004. For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London B* (Special Issue on Law and the Brain) 359: 1775–1785.
10. Wittgenstein, L. 1953/2001. *Philosophical investigations: 50th anniversary edition*. Oxford: Blackwell. Translated by G.E.M. Anscombe.
11. Greely, H. 2009. Neuroscience-based lie detection: The need for regulation. In *Using imaging to identify deceit: Scientific and ethical questions*, ed. E. Bizzi et al. Cambridge: American Academy of ARts and Sciences.
12. Kozel, F., K. Johnson, Q. Mu, E. Grenesko, S. Laken, and M. George. 2005. Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry* 58: 605.
13. Langleben, D., L. Schroeder, J. Maldjian, R.C. Gur, S. McDonald, J.D. Ragland, et al. 2002. Brain activity during

simulated deception: An event-related functional magnetic resonance study. *Neuroimage* 15: 727.
14. Davatzikos, C., K. Ruparel, Y. Fan, D.G. Shen, M. Acharyya, J.W. Loughead, R.C. Gur, and D.D. Langleben. 2005. Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. *Neuroimage* 28: 663.
15. Ganis, G., S.M. Kosslyn, S. Stose, W.L. Thompson, and D.A. Yurgelun-Todd. 2003. Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex* 13: 830.
16. Langleben, D., J. Loughead, W. Bilker, K. Ruparel, A.R. Childress, S. Busch, and R.C. Gur. 2005. Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping* 26: 262.
17. Lee, T.M.C., H. Liu, L. Tan, C. Chan, S. Mahankali, C. Feng, J. Hou, P. Fox, and J. Gao. 2002. Lie detection by functional magnetic resonance imaging. *Human Brain Mapping* 15: 157.
18. Mohamed, F.B., et al. 2006. Brain mapping of deception and truth telling about an ecologically valid situation: Function MR imaging and polygraph investigation—initial experience. *Radiology* 238: 679.
19. Nunez, J.M., B.J. Casey, T. Egner, T. Hare, and J. Hirsch. 2005. Intentional false responding shares neural substrates with response conflict and cognitive control. *Neuroimage* 25: 267.
20. Spence, S.A., T. Farrow, A. Herford, I. Wilkinson, Y. Zheng, and W.R. Peter. 2001. Behavioral and functional anatomical correlates of deception in humans. *Brain Imaging Neuroreport* 2849.
21. Greene, J.D., and J.M. Paxton. 2009. Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences of the United States of America* 106(30): 12506–12511.
22. Farah, M.J., and P.R. Wolpe. 2004. Monitoring and manipulating brain function: New neuroscience technologies and their ethical implications. The Hastings Center Report, Vol. 34.
23. Garland, B., and P. Glimcher. 2006. Cognitive neuroscience and the law. *Current Opinion in Neurobiology* 16(2): 130–134.
24. Greely, H. T., and J. Illes. 2007. Neuroscience-based lie detection: The urgent need for regulation. *American Journal of Law and Medicine* 33(2–3): 377–431.
25. Kanwisher, N. 2009. The use of fMRI in lie detection: What has been shown and what has not. In *Using imaging to identify deceit: Scientific and ethical questions*, ed. Emilio Bizzi et al., 7–13. Cambridge: American Academy of ARts and Sciences.
26. Kittay, L. 2007. Admissibility of fMRI lie detection—the cultural bias against mind reading devices. *Brooklyn Law Review* 72: 1351.
27. Langleben, D. 2008. Detection of deception with fMRI: Are we there yet? *Legal and Criminological Psychology* 13(1): 1–9.
28. Moreno, J.A. 2009. The future of neuroimaged lie detection and the law. *Akron Law Review* 42: 717–734.
29. Phelps, E. 2009. Lying outside the laboratory: The impact of imagery and emotion on the neural circuitry of lie detection. In *Using imaging to identify deceit: Scientific and ethical questions*, ed. Emilio Bizzi et al., 14–22. Cambridge: American Academy of ARts and Sciences.
30. Rakoff, J.S. 2009. Lie detection in the courts: The vain search for the magic bullet. In *Using imaging to identify deceit: Scientific and ethical questions*, ed. Emilio Bizzi et al., 40–45. Cambridge: American Academy of ARts and Sciences.
31. Schauer, F. 2010. Can bad science be good evidence: Lie detection, neuroscience, and the mistaken conflation of legal and scientific norms. *Cornell Law Review*, forthcoming.
32. Sinnott-Armstrong, W., et al. 2009. Neural lie detection in courts. In *Using imaging to identify deceit: Scientific and ethical questions*, ed. Emilio Bizzi et al., 35–39. Cambridge: American Academy of ARts and Sciences.
33. Spence, S. 2008. Playing devil's advocate: The case against fMRI lie detection. *Legal and Criminological Psychology* 13(1): 11–25.
34. Henig, R.M. 2006. Looking for the lie. *New York Times Magazine*. February 5, 2006.
35. Narayan, A. 2009. The fMRI brain scan: A better lie detector? TIME Magazine.
36. Silberman, S. 2006. Don't even think about lying. *Wired* 14.01.
37. Wittgenstein, L. 1965. *Preliminary studies for the "philosophical investigations", generally known as the blue and brown books*. New York: Harper & Row.
38. Caraway, C. 1986. Criteria and conceptual change in Wittgenstein's later philosophy. *Metaphilosophy* 17(2): 162–171.
39. Klagge, J. 1989. Wittgenstein and neuroscience. *Synthese* 78(3): 319–343.
40. Koethe, J.L. 1977. The role of criteria in Wittgenstein's later philosophy. *Canadian Journal of Philosophy* 7(3): 601–622.
41. Albritton, R. 1959. On Wittgenstein's use of the term 'criterion'. *Journal of Philosophy* 56: 845–857.
42. Caraway, C. 1984. Criteria and circumstances. *The Southern Journal of Philosophy* XXII(3): 307–316.
43. Chihara, C.S., and J.A. Fodor. 1965. Operationalism and ordinary language: A critique of Wittgenstein. *American Philosophical Quarterly* 2: 281–295.
44. Garver, N. 1962. Wittgenstein on criteria. In *Knowledge and experience*, ed. C.D. Rollins, 55–87. Pittsburgh: University of Pittsburgh Press.
45. Hollinger, R. 1974. Natural kinds, family resemblances, and conceptual change. *The Personalist* 55: 323–332.
46. Kenny, A. 1967. Criterion. *Encyclopedia of Philosophy* 260–261.
47. Malcolm, N. 1957. *Dreaming*. London: Routledge and Kegan Paul.
48. Putnam, H. 1962. Dreaming and 'depth grammar'. In *Analytical philosophy*, ed. R.J. Butler, 211–235. Oxford: Basil Blackwell.
49. Scriven, M. 1959. The logic of criteria. *Journal of Philosophy* 56: 857–868.
50. Lycan, W.G. 1971. Noninductive evidence: Recent work on Wittgenstein's "criteria". *American Philosophical Quarterly* 8(2): 109–125.
51. Knobe, J. 2007. Experimental philosophy and philosophical significance. *Philosophical Explorations* 10(2): 119–121.
52. Knobe, J., and S. Nichols. 2008. An experimental philosophy manifesto. In *Experimental philosophy*, ed. J. Knobe and S. Nichols, 1–14. New York: Oxford University Press.
53. Nadelhoffer, T., and E. Nahmias. 2007. The past and future of experimental philosophy. *Philosophical Explorations* 10 (2): 123–149.