What Brains Won’t Tell Us About the Mind: A Critique of the Neurobiological Argument Against Representational Nativism

RICHARD SAMUELS

Abstract: In their recent and influential book *Rethinking Innateness*, Jeffrey Elman and his co-authors argue that evidence from neurobiology provides us with grounds to reject representational nativism (RN). I argue that Elman et al.’s argument fails because it makes a series of unwarranted assumptions about RN and about the extent to which neurobiological data constrain claims about the innateness of mental representations. Moreover, I briefly discuss how we ought to understand RN and argue that on two prima facie plausible approaches, far from refuting nativism, the evidence from neurobiology may not even be relevant to the question of whether or not RN is true.

Introduction

Over the past two thousand years the doctrine of innate ideas has had some notable proponents. Plato, Descartes, Leibniz, Reid and, in the twentieth century, Noam Chomsky have all developed and defended conceptions of the human mind in which innate ideas or innate representations play a central role. Nor, of course, has the doctrine been without its detractors. In the seventeenth and eighteenth centuries the British empiricists predicated their philosophical systems on the rejection of innate ideas. And, since that time, empiricist psychologists and philosophers of mind have proceeded under the assumption that positing innate ideas is, at best, unnecessary and, at worst, mystical and unscientific (Cowie, forthcoming). In spite of empiricist scruples, however, recent cognitive science has witnessed a resurgence of
interest in the doctrine of innate ideas—or representational nativism as it is now commonly called—the view that the human mind contains innately specified representations. Largely under the influence of Chomsky’s work in linguistics, representational nativism has become an important theoretical option in cognitive science and has been invoked in order to explain a wide range of phenomena. For example, it has been argued that we possess innate representations of the principles of grammar (Pinker, 1994), physics (Carey and Spelke, 1994), mathematics (Wynn, 1992) and psychology (Leslie, 1994). Despite its theoretical utility, however, the positing of innate representations remains contentious.

In this paper, I focus on one of the most recent and intriguing attempts to undermine representational nativism (RN). In an influential and widely cited work entitled *Rethinking Innateness*, Jeffrey Elman and his co-authors argue that evidence from neurobiology provides us with good reason to conclude that ‘the case for innate representations does not look very good’ and that RN should be rejected (Elman et al., p. 27). Clearly, if they are correct, the implications are profound. Not only would Elman et al. have succeeded in settling the centuries-old dispute over innate ideas but their argument would also require the wholesale rejection of a range of widely accepted theories in cognitive science. In what follows, however, I will argue that Elman et al.’s neurobiological argument fails to undermine RN. ¹

In order to do this, I must first say what the neurobiological argument is. In section 1, I provide a brief outline of Elman et al.’s argument. Then, in section 2, I make a few preliminary comments about how Elman et al. characterize the notion of innateness. In the remainder of the paper I show that the neurobiological argument fails to undermine RN because it makes a series of unwarranted assumptions about nativism and about the extent to which neurobiological data constrain claims about the innateness of mental representations. In section 3, I argue that Elman et al. incorrectly assume that the innately specified properties of a piece of cortical tissue must be intrinsic properties of the tissue. In section 4, I argue that even if we assume (contrary to fact) that innate properties must be intrinsic, the neurobiological argument still fails because it depends upon overly strong assumptions about the inflexibility or invariance of innately specified structures. Next, in section 5, I argue that even if (contrary to fact) both of the above assumptions are true, the neurobiological argument still fails because it overestimates the extent to which neurobiological evidence constrains claims about the innateness of mental representations. In section 6, I conclude by briefly considering the issue of how we ought to understand debates about representational nativism and argue that on two prima facie plausible approaches, far from refuting RN, the evidence from invasive, developmental, neurobiolog-

¹ Steven Quartz and Terrence Sejnowski (among others) have developed very similar arguments to those formulated by Elman et al. See, for example, Quartz and Sejnowski, 1994. As far as I can see, my criticisms also apply mutatis mutandis to Quartz and Sejnowski’s arguments.
ical studies may well not even be relevant to the question of whether or not RN is true.

1. The Neurobiological Argument

Let me start by providing a brief statement of the neurobiological argument against representational nativism. I’ll first give a general outline of the argument. Then, I’ll describe some of the experimental data in more detail.

1.1 General Outline

According to Elman et al., ‘evidence has been mounting against the notion of innate domain-specific microcircuitry as a viable account of cortical development, i.e. against what we call “representational nativism”’ (p. 26). The evidence in question comes from a variety of studies on cortical plasticity:

In a number of recent studies with vertebrate animals, investigators have changed the nature of the input received by a specific area of cortex, either by transplanting plugs of fetal cortex from one area to another (e.g. somatosensory to visual or vice versa, O’Leary, 1993; O’Leary and Stanfield, 1989), by radically altering the nature of the input by deforming the sensory surface (Friedlander, Martin and Wassenhove-McCarthy, 1991; Killackey et al., 1994) or by redirecting inputs from their intended target to an unexpected area (e.g. redirecting visual inputs to auditory cortex (Frost, 1983, 1990; Pallas and Sur, 1993)). (p. 26)

In section 1.2, I will discuss some of these studies in greater detail. For the moment, however, the key point to note is that, according to Elman et al., the outcome of these studies poses a problem for representational nativism. Elman et al. appear to believe that if RN were true, then the representational properties of foetal cortical tissue—i.e. what the pieces of cortical tissue represent—would not be significantly altered as a result of the above kinds of experimental manipulations. But this is not what occurs:

Surprisingly, under these aberrant conditions, the fetal cortex takes on neuroanatomical and physiological properties that are appropriate for the information received (‘When in Rome, do as the Romans do . . .’), and quite different from the properties that would have emerged if the default inputs for that region had occurred. This suggests that cortex has far more representational plasticity than previously believed. (pp. 26–7)

Elman et al. conclude that ‘the cortex appears to be an organ of plasticity’—an organ whose representational properties are highly flexible in response
to environmental change (p. 315). From this they infer that while ‘one cannot entirely rule out the possibility that neurons are born “knowing” what kinds of representations they are destined to take on . . . right now the case for innate representations does not look very good’ (pp. 26–7).

1.2 Some Experimental Data

In order to gain a deeper understanding of the argument, it is useful to consider in more detail some of the experiments cited by Elman et al. I’ll discuss two kinds of experiments: O’Leary et al.’s transplant studies and Sur et al.’s ‘rewiring’ experiments.

Transplant Studies In a series of experiments, O’Leary and his colleagues successfully transplanted pieces of foetal cortical tissue from one region of the newborn rodent cortex to another (O’Leary and Stanfield, 1989; Stanfield and O’Leary, 1985). What they discovered is that the transplanted tissue takes on the structural and functional properties of its new location as opposed to maintaining the structure/function of its development origins. Moreover, the structure/function of the transplanted tissue appears to be determined, in part, by the character of the sensory inputs it receives. So, for example, in one experiment, tissue from the visual cortex was transplanted to a part of the somatosensory region of the rodent cortex that normally develops a distinctive ‘barrel field’ structure. When innervated by thalamic afferents (i.e. sensory inputs that are mediated via thalamic structures), the transplanted tissue developed a structure very similar to the barrel fields that are observed in the normally developing rodent somatosensory region as opposed to developing the structure that is characteristic of the rodent visual cortex (Schlaggar and O’Leary, 1991). Furthermore, it also developed the pattern of input and output projections appropriate to its new location. Thus, as a result of the transplant, both the internal structure of the transplanted visual cortical tissue and the pattern of connections it had to other parts of the cortex were altered.

‘Rewiring’ Experiments A number of ‘rewiring’ experiments have been performed in order to determine the extent to which the mammalian cortex is plastic (Sur et al., 1990). The basic strategy is to induce projections from the retina to project to the auditory thalamic areas and then onto the auditory cortex instead of taking their normal route to the visual cortex via the lateral geniculate nucleus (LGN). This is achieved by (a) selectively lesioning parts of the visual cortex and LGN so that exuberant (i.e. ‘extra’) connections that would ordinarily be eliminated continue to grow, and (b) making lesions that prevent auditory inputs from innervating their normal thalamic targets:

Under these pathological conditions retinal projections will reroute (i.e. the exuberant connections will remain and continue their journey) to innervate the medial geniculate nucleus (MGN), which
is of course the way station that normally handles auditory inputs. Projections from the MGN then project to the auditory cortex, in the normal fashion. (Elman et al., p. 273)

In short: it is possible to ‘rewire’ the cortex so that retinal inputs are processed by the auditory cortex instead of the visual cortex. Under these conditions it turns out that the auditory cortex becomes responsive to visual inputs and develops many of the representational properties that are characteristic of the normally developing visual cortex—e.g. ‘visual maps’. Thus it would seem that, when parts of the cortex are suitably ‘rewired’, their representational properties can be radically altered.

There are some obvious worries one might have about whether the above evidence really provides strong support for the claim that the human cortex is an ‘organ of plasticity’. So, for example, it is unclear how strong the inference is from the experimental evidence cited by Elman et al. to the conclusion that the whole cortex is highly plastic. Moreover, one might be concerned about the inference from animal studies to conclusions about human beings. In what follows, however, I will ignore these worries and assume for the sake of argument that the evidence for viewing the human cortex as an ‘organ of plasticity’ is compelling. What I propose to argue is that even if we accept this conclusion, the argument still fails to provide us with any reason to reject RN.

2. Definitional Manoeuvres

One way to facilitate the refutation of RN is to define ‘innate’ in such a way that RN is easily refutable. Elman et al. often appear to be guilty of attempting to win by employing this sort of manoeuvre. For if we take their characterization of ‘innate’ seriously, then we are left with a straw-man nativism that is easily refuted but accepted by no one. In this section, I briefly discuss and criticize the definition of ‘innate’ adopted by Elman et al.

After a preliminary discussion of the various ways in which theorists have characterized the term ‘innate’, Elman et al. opt for the following definition:

(1) A characteristic (e.g. ‘an aspect of brain structure, cognition or

---

2 This is particularly troubling because (for obvious reasons) none of the neurobiological experimental findings cited by Elman et al. concerns such ‘higher’ cognitive abilities as language processing or folk psychological reasoning. But it is for precisely these sorts of higher abilities that contemporary representational nativists are most keen to stake their claim.

3 For a discussion of such inferences see Preuss, 1995.

4 Elman et al. consider three different characterizations of what ‘innate’ means. None of these accounts is satisfactory but due to limitations of space I shall not consider them all.
behavior’ is innate just in case it is ‘the product of interactions internal to the organism’ (p. 23).

It is clear from the context that they mean that innate characteristics are solely the product of interactions internal to the organism. Thus, according to (1), a characteristic cannot be innate if it is, in part, the product of interactions that the organism has with the environment.

If this is what we mean by ‘innate’, then the experiments cited by Elman et al. may well show that RN is false. After all, what the experiments suggest is that the development of representational structure in specific cortical regions is, at least in part, dependent on the character of the environmentally derived stimuli that an organism receives. But, then again, if this is what ‘innate’ means, the sophisticated neurobiological experiments are not needed in order to refute RN. This is because it is very plausible to claim that all human characteristics depend on at least some interactions with things external to the organism (Block, 1981, p. 281). A foetus does not develop legs, arms and skin without exchanging oxygen, water and nutrients with its mother. A neonate does not develop teeth and hair without breathing, drinking and eating. But all of this involves interaction with the environment external to the organism. So, if ‘innate’ means what Elman et al. appear to be suggesting, then RN is clearly false, but so too is nativism about arms, legs and teeth. Yet this is surely not a problem for the representational nativist. Rather, it is a problem for the way in which Elman et al. chose to characterize the notion of innateness. Nativists do not claim that innate characteristics must be acquired without any interactions between the organism and the environment. They do not deny, for example, that the development of innately specified concepts (or legs or hair) requires the intake of environmentally derived nutrients. Consequently, if the success of the neurobiological argument requires that innate characteristics are defined as those which are solely the product of interactions internal to the organism, then it is an argument that only succeeds in destroying a straw man.

Yet perhaps there is a way of characterizing the notion of ‘interaction’ employed in (1) so as to avoid the above criticism. Elman et al.’s discussion appears to incorporate such a proposal. At one point they suggest that what’s ruled out by nativism is the utilization of information that’s derived from ‘outside the organism’ (p. 22). Thus nativism about a characteristic C does not preclude all organism-environment interactions but merely interactions that involve the transmission of information from the environment to the organism. In short:

(2) A characteristic C is innate just in case it develops without ‘recourse to information from outside the organism’ (p. 22).

This is an improvement on the earlier proposal. Nonetheless, it is still subject to a number of serious problems. First, this characterization of ‘innate’ depends upon an unspecified notion of information. And what counts as an
innate characteristic will vary depending on what we count as information. So, for example, according to some notions of information, maternal hormones and environmentally derived nutrients might be said to convey information. In which case we are back to the problem discussed above, i.e. the characterization of ‘innate’ turns RN into a straw man. So, in order to make (2) plausible as a characterization of innateness, we need to characterize ‘information’ in such a way that only a highly restricted class of environment/organism interactions involve the transfer of information.

A second problem is that, as a matter of fact, nativists typically do not deny that environmentally derived information is relevant to the development of innate representations. Indeed they frequently insist that information about the environment, conveyed in the form of sensory stimuli, play a central role in the development of innate representations, but only by triggering the representation, rather than by playing the role of evidence in a learning process (Chomsky, 1980, p. 33; Fodor, 1981, pp. 274–5). Precisely what ‘triggering’ is supposed to be is an interesting question and one that nativists have been insufficiently clear in addressing. Nevertheless given that nativists agree with non-nativists that environmentally derived information is causally implicated in the acquisition of representations, one obviously cannot refute RN merely by pointing out that environmentally derived information is causally implicated in the acquisition of representations. Given that this is so, there are two points that we ought to keep in mind when discussing the neurobiological argument: First, (2) above does not accurately capture what nativists typically mean by ‘innate’. Second, since learning and triggering both involve environmentally derived information, the neurobiological argument cannot undermine RN merely by showing that exchanges of information from the environment to the organism are causally implicated in the acquisition of representations.

5 Here is what Jerry Fodor has to say on the matter: ‘Both [nativist (AKA Rationalist) and empiricist] theories assume that the availability of . . . concepts is contingent upon the activation of the sensorium. As Descartes frequently emphasized, it is not part of the Rationalist program to claim that your repertoire of primitive concepts is independent of the character of your experience. On the contrary, just as there are presumably triggering experiences that are nomologically necessary for the acquisition of sensory-primitive concepts like red, so are there presumably triggering experiences that are nomologically necessary for the acquisition of non-sensory primitive concepts like triangle. In either case, if you don’t encounter the trigger, you don’t get the concept’ (Fodor, 1981, p. 275).

6 Fodor characterizes triggering as a process for making representations (or concepts) available that is ‘brute-causal’ as opposed to ‘rational’. According to Fodor, a rational process of concept acquisition is one in which ‘the experiences which eventuate in the availability of . . . a concept are held to bear a confirmation relation to some hypothesis which specifies the internal structure of the concept’ (Fodor, 1981, p. 272). A brute-causal process for making concepts or representations available is one that is not rational. Notice, however, that according to this characterization of ‘triggering’, triggering is not sufficient for innateness. After all, if I acquired knowledge of Latin by taking a Latin pill, the process of acquisition would be nonrational, hence, brute-causal, hence, triggered. But my knowledge of Latin surely wouldn’t be innate.
3. Innate and Intrinsic Properties

The evidence cited by Elman et al. for rejecting representational nativism is of the following form: if one alters the relationship between afferent inputs and pieces of neural tissue (e.g. by transplanting the tissue or redirecting the inputs), then the representational properties of the tissue change. On the basis of this kind of evidence, Elman et al. conclude that since the representational properties of the neural tissue do change as a result of experimental manipulations, these properties are not innately specified. But this sort of evidence against RN is unconvincing since it confounds issues about the innately specified properties of cortical tissue with issues about the intrinsic properties of cortical tissue. Very roughly, the intrinsic properties of an object are those properties that it has independently of the relationships it bears to other things. An extrinsic property is simply one that is not intrinsic.\(^7\) Now, if representational properties are extrinsic, then the relations that a piece of cortical tissue bears to other parts of the brain (and/or the world at large) will be partially constitutive of the representational properties it possesses. In which case, the studies cited by Elman et al. provide us with no good reason to reject RN. For if representational properties are extrinsic, then:

(a) It may well be innately specified that a piece of cortical tissue T possess a representational property P (where its relations to other parts of the brain and/or the world at large are partially constitutive of P), and
(b) T might no longer possess P when it is transplanted to a different neural region, because
(c) T’s relations to other parts of the brain (and/or the extra-neural world) would have been altered.

In short: if representational properties are extrinsic (or relational), then it should come as no surprise that alterations in the relations of neural structures to other parts of the brain (and/or the world at large) would alter the representational properties of those structures. After all, one may well be altering relations that are partially constitutive of the representational properties.

The above observation would be of little use as a response to the neurobiological argument if either (a) extrinsic properties could not be innate, or else (b) it was implausible to claim that representational properties are extrinsic. As a matter of fact, however, there is no reason to suppose that properties cannot be both innately specified and extrinsic. So, for example, it is presumably an innately specified property (if anything is) of the human

---

\(^7\) See Lewis, 1983, for a discussion of the distinction between intrinsic and extrinsic properties.
head that it grows so as to be connected to the neck. But being connected to a neck is a relational and, hence, extrinsic property of the head. So, extrinsic properties can be innately specified. Moreover, it is not implausible to claim that representational properties are extrinsic. Indeed according to virtually all extant theories of meaning or content, representational properties are extrinsic properties of the representation. So, according to these theories, it should hardly come as a surprise that if we alter the relations that a neural object bears to the rest of the brain and the world at large (e.g. by transplanting it), then it will represent something different from what it had previously represented. By way of illustration, consider perhaps the two most influential, contemporary approaches to meaning and content: information-based semantics and conceptual role semantics.

According to information-based semantics, a mental (or neural) representation has its content or meaning in virtue of certain relations that obtain between tokens of the representation and properties (or states of affairs) of the world. Thus, on such an account, content is an extrinsic property par excellence. Moreover, it is an extrinsic property that should be expected to be sensitive to the kinds of experimental manipulations discussed by Elman et al. This is because it is ordinarily assumed by informational semanticists that the relations to the world in virtue of which mental (or neural) representations have their content are mediated via the senses. For example, a representation with the content red, will have this content, in part, because the visual system enables it to bear the appropriate relation to red things in the environment. But if this is so, then experimental manipulations which alter the relationship between a representation and the world (such as transplanting the neural representation from the visual cortex to the somatosensory cortex), can also be expected to alter the content of the representation.

Content also turns out to be an extrinsic property of representations according to conceptual role semantics. On this approach (very roughly), a representation has its content in virtue of the causal relations that it bears to other representations. So, for example, the representation dog means dog in virtue of its relations to other representations—e.g. the relation between dog and the representation mammal. Moreover, if we adopt Elman et al.’s assumption that some neural structures are representations (or, at any rate,

---

8 See Stich and Warfield, 1994, for a recent anthology of papers on contemporary theories of content.
9 So, for example, on one very simple version of the theory, a given representation R means (has the content) dog because tokens of the representation are reliably caused by instances of doghood in the environment.
10 I adopt the following notational conventions: mentioned (mental) representations are indicated by capitals and mentioned properties are indicated by italics. I also use italics for emphasis.
implement representations), then it should once more come as no surprise that altering the inputs to a neural structure or transplanting it from one cortical region to another should affect the content of the representation. After all, such experimental manipulations may radically alter the causal relations that the neural structure bears to other representations that are implemented in the brain and, hence, alter the content of the representation.

Let me summarize. If what I have said in the last few paragraphs is correct, then it is plausible to think that representational properties are extrinsic properties of pieces of cortical tissue and that the sorts of experimental manipulations discussed by Elman et al. should be expected to affect the content of neural/mental representations. Moreover, as we have already seen, there is no problem with the claim that innately specified properties of an object are extrinsic. In which case, the evidence cited by Elman et al. is perfectly consistent with the claim that there are innately specified representational properties of mental and neural structures.

4. Innateness and Invariance

Let us now suppose, for the sake of argument, that all the innate properties of an object are intrinsic properties of that thing. So, by assumption, if a representational property \( P \) (e.g. representing \( \text{red} \)) of a piece of neural tissue is innately specified, then \( P \) is an intrinsic property of the neural tissue. Would the neurobiological argument then succeed in showing that RN is false? Once again the answer is ‘No’. One serious problem is that the argument depends upon the following implausible principle:

*Principle of Invariance* The innately specified (representational) properties of a piece of cortical tissue \( T \) are invariant under alterations in \( T \)’s location within the brain and alterations in the afferent inputs to \( T \).

The rough idea is that if a property of a piece of neural tissue is innate—if it is innately specified, for example, that the tissue represent the color \( \text{red} \)—then the tissue ought to be inflexible or unplastic with respect to that property. Specifically, experiments which involve transplanting the tissue from

---

11 Elman et al. are far from clear about what precisely the relationship is between representations and neural structures. Sometimes they speak as if the relation is identity and sometimes they speak as if the relationship is a weaker one—such as implementation or realization. I discuss this issue in more detail in section 5.

12 Obviously, this principle needs to be hedged in various ways in order to be rendered (even superficially) plausible. So, for example, presumably no one insists that cortical tissue must hold onto its innate properties even when it is severely damaged by experimental manipulation. So, we ought to add the following caveat to the above principle: ‘... unless seriously damaged’. Since my criticisms do not turn on the precise formulation of the principle, however, I will not bother to provide a precise formulation.
one brain region to another, or which involve changing the afferent inputs to the tissue, ought not to affect whether or not the tissue possesses the innately specified property. Elman et al. never explicitly state the above principle. Nevertheless they do hint at it, for example, when characterizing RN as committed to the claim that innate representations are "‘hardwired’ in advance" (p. 25). Furthermore, it should be clear that the neurobiological argument has no chance of working unless we make such an assumption. After all, if the innately specified properties of neural tissue need not conform to the Principle of Invariance, then the fact that the properties of neural tissue alter as a result of experimental manipulation would not count against RN.

How plausible is the Principle of Invariance? The answer, I maintain, is that it is not very plausible at all. If the Principle of Invariance were plausible, then presumably nativists about noncortical features of the body ought to be committed to analogous principles. For the proponent of RN surely does not want to claim that innate representations are somehow more innate than skin, hearts, teeth and the like. But it would presumably be singularly implausible to conclude that a certain group of cells C (in a normally developing organism) is not innately specified to become, say, teeth, on the grounds that if C were transplanted to a different location in the organism early enough in development, then it would develop into (say) part of the belly region. In other words, the claim that C is innately specified to develop into teeth (or hearts or skin etc.) does not entail that C is subject to a principle of invariance.

Nor is the above example merely a fanciful thought experiment. Some classic studies from developmental biology provide us with genuine cases where pieces of tissue that are transplanted from one (noncortical) region of the developing organism to another will behave according to their new location and not their developmental origin. So, for example, Spemann and his colleagues transplanted epidermis tissue from the belly region of developing amphibians to the area fated to become the mouth region, with the result that the transplant formed the mouth and teeth according to its new location (Spemann, 1938). But on any reading of ‘innately specified’ that is strong enough for any (sane) representational nativist, we surely would not want to conclude that it is not innately specified that the transplanted cells (in the normally developing amphibian) become part of the belly region. That is, claims about the innate specification of noncortical tissue do not entail that the tissue is subject to a principle of invariance. Why, then, should be proponent of RN be committed to such a principle? As far as I can see there is no reason.

---

13 See Muller, 1997, and Gilbert, 1994, for other examples of transplant studies in developmental biology.
5. The Inference from Neural Plasticity to the Absence of Innate Mental Representations

Let us suppose, however, that the innately specified properties of neural tissue must satisfy the Principle of Invariance. Would the experiments cited by Elman et al. now show that representational nativists are wrong? Yet again, the answer is ‘No’. In order to make this point, let me start by distinguishing between two versions of RN; what I call tissue nativism and organism nativism. Tissue nativism is a thesis about the innately specified properties of pieces of cortical tissue. According to tissue nativism, there are representational properties that specific pieces of cortical tissue are innately specified to possess. To express the point in a slightly different way, according to tissue nativism, specific pieces of cortical tissue are innately specified to encode certain representations. So, for example, it might be innately specified that a specific cluster of neurons N in the prefrontal lobe encode the representation belief.

In contrast to tissue nativism, organism nativism is a thesis about whole organisms and the mental representations that they possess. According to organism nativism, it is innately specified that organisms possess certain mental representations. What it claims is that people—or more generally organisms—possess innately specified mental representations that encode for various forms of domain specific knowledge. So, for example, it might be innately specified that human beings possess such mental representations as belief, one and object.

With the above distinction in hand, let’s consider the question of what can be inferred from the conjunction of the Principle of Invariance and the experimental data cited by Elman et al. Suppose that the Principle of Invariance is true. Moreover, suppose, for the sake of argument, that the experimental data cited by Elman et al. provide support for the idea that the human cortex is an ‘organ of plasticity’—specifically, that all the representational properties of any piece of cortical tissue can be altered as a result of experimental manipulations of the sort Elman et al. describe. Then we may infer the following:

\[ \neg \text{TN} \] It is not innately specified that any particular piece of neural tissue has any specific representational property P.

Now, if this is true, then it follows that tissue nativism is false. After all, tissue nativism claims that it is innately specified that specific pieces of cortical tissue code for certain representations. So, if (contrary to fact) we

---

14 More precisely, organism nativists claim that: (a) where types of mental representations are individuated by their contents, (b) there are some types of mental representations (e.g. the type belief or red), such that (c) it’s innately specified that we all possess tokens of that type of representation.
accept both the principle of invariance and the thesis that the cortex is highly
plastic, then we appear to have good reason to reject tissue nativism.
This conclusion, however, is unlikely to worry the representational nativist. This is because it is clear that the version of representational nativism endorsed by contemporary theorists—such as Chomsky, Fodor, Leslie, Gopnik, Spelke and Carey—is organism nativism and not tissue nativism. That is, contemporary theorists who defend nativism about representations are concerned with claims about what innate mental representations people (and other organisms) possess and not claims about the properties of specific pieces of neural tissue. So, for example, Carey and Spelke (1994) argue that ‘human reasoning is guided by a collection of innate domain-specific systems of knowledge’ where ‘each system is characterized by a set of core principles that define the entities covered by the domain’ (Carey and Spelke, 1994, p. 169). Notice, however, that this is a claim about organisms—specifically, human beings—and not a claim about the innate properties of specific pieces of cortical tissue. By contrast, (¬TN) expresses a claim about the properties of neural tissue. It expresses a fact (if indeed it is a fact) about our neurobiology: a neural-level claim and not a psychological one. And clearly we cannot directly infer from (¬TN) that organism nativism is false—that organisms possess no innately specified mental representations.
So, it would appear that even if Elman et al. provide us with reason to reject tissue nativism, this ought not to be a source of immediate concern for contemporary representational nativists, since they defend organism nativism. But what is the relationship between these two forms of representational nativism? In particular, if tissue nativism is false, do we also have reason to reject organism nativism? I will now argue that the answer is ‘No’.
Presumably, if the assumption that tissue nativism is false were to provide us with reason to reject organism nativism, then this would be because of the relationship between neural structures (states or processes) and the mental structures (states or processes) of organisms. A common assumption in contemporary cognitive science about the relationship between the mental and the neurophysical is that mental states (e.g. beliefs and desires) and mental entities (such as concepts, mental representations and inferential mechanisms) are multiply realizable; that tokens of the same type of psychological entity (state or processes) can be realized by different kinds of neural entity (state or process). The point is perhaps clearest in the case of belief, where (now familiar) thought experiments have convinced most theorists that people can possess the same (type of) belief while being in different neural states. Imagine, for example, that you and I both believe that Paris is the capital of France. Since your belief has the same content as mine, it follows that if we individuate belief types by their contents, then you and I

15 Similar claims abound in the developmental literature. See, for example, Spelke, 1995; Gopnik and Meltzoff, 1997, and Leslie, 1994.
16 For more detailed discussions of multiple realizability see Block, 1980, and McDonald and McDonald, 1995.
What Brains Won’t Tell Us About the Mind

have (token) beliefs of the same type. Nevertheless, it is surely logically possible that what is going on in your brain when you token the belief that Paris is the capital of France is not the same thing that’s going on in my brain when I token that belief. An intrapersonal version of this thought experiment is also plausible. Suppose that at time t1, I believe that Paris is the capital of France, at time t2, I suffer focal brain damage which results in my losing the neural structures that had previously realized this belief, and at time t3, I once more acquire the belief that Paris is the capital of France. In this thought experiment I will, at different times, have tokens of the same belief type, but the two different tokens will be realized by different neural structures.

Suppose that mental entities are multiply realizable. Then it is clear that tissue nativism may be false while organism nativism is true. One way of illustrating this claim is to consider cases where it is innately specified that some cortical structure realizes a given mental representation R, but it depends on contextual factors which specific structure is the realizer of R. So, for example, it may be that in context C1, neural structure N1 realizes R; whereas in C2, N2 realizes R, and so on. There are known to be such cases of developmental branching. In primates, for instance, a region of the temporal cortex, TE, normally plays a critical role in visual object recognition. But when TE is lesioned early in development, certain parietal regions of the cortex—STP, PG and TF—are recruited in order to perform visual object recognition tasks (Elman et al., p. 278–9). As Elman et al. point out, this example suggests that ‘Higher functions like visual object recognition can be organized in a variety of ways’ (p. 279). Now, Elman et al. clearly think that this provides evidence against RN. But in fact it is entirely compatible with everything we know about the brain and its development that the above case of developmental branching is innately specified. It might be innately specified that when TE is damaged, STP, PG and TF are recruited to perform the task of visual object recognition. Moreover, if this is the case, then clearly tissue nativism might be false while organism nativism is true. For in such a case it might be innately specified that some region of the cortex subserve visual object recognition, even though it is not innately specified that any specific region—e.g. TE—be responsible for the task.

But suppose we were to reject the assumption that mental structures are multiply realizable by neurophysiological structures, and instead adopt some stronger thesis about the relationship between mental entities and neural entities. In particular, suppose we were to endorse a type-type identity

---

17 It seems likely that similar cases occur in human infancy. So, for example, there are instances where, in spite of damage to the left hemispherical regions normally employed in language processing, people develop normal linguistic abilities (and within the normal time-frame). Among other things, this sort of example suggests that human beings can be in the same intentional state—e.g. knowing English or possessing the move-alpha rule—in spite of the fact that these intentional states are subserved by different underlying neural structures.
thesis according to which every type of mental entity is identical with some type of neural entity—e.g. that being pain = C-fibre activity, being the concept \textit{red} = 30 MHz activity in the frontal cortex and so on. It is widely agreed that such a thesis is implausible. Suppose, however, that it is true. Would this render organism nativism untenable? Yet again the answer is ‘No’. In order to make the point, let me start by considering an analogous case that doesn’t involve the cortex: skin regeneration. The surface of the human body is covered by skin. Indeed it is surely innately specified that our bodily parts are thus covered. But we can distinguish between two distinct versions of nativism about skin that are close analogues of organism and tissue nativism:

1. \textbf{Organism/skin nativism:} it is innately specified that the human body is covered by skin.
2. \textbf{Tissue/skin nativism:} it is innately specified that specific pieces of skin cover specific pieces of the body—e.g. skin piece \( S_1 \) covers my forehead, \( S_2 \) covers my left knee and so on.

Now if I were to graze my knee, a sequence of physiological repair activities would become operative and result in the regeneration of new skin. When the sequence of activities is complete, the result is a new piece of skin that is, to all intents and purposes, type-identical to the old one. This process of regeneration is surely innately specified. (If it is not, then what is?) But notice, it need not be that tissue/skin nativism is true: it need not be innately specified that \textit{this} particular new piece of skin cover my knee. Instead organism/skin nativism need only be true: it need only be the case that \textit{some} piece—any piece—of skin cover the knee.

Now apply this healing analogy to the cortical case. Suppose that, at time \( t_1 \), there is a piece of cortex \( N_1 \) that encodes the representation \textit{red}. At time \( t_2 \), \( N_1 \) is moved or rewired, as a result of which the representational properties of \( N_1 \) change—it no longer encodes the representation \textit{red}. Given the principle of invariance, it follows that \( N_1 \) is not innately specified to encode \textit{red}. So, tissue nativism is not true of \( N_1 \). Nevertheless, let us further suppose that subsequent to time \( t_2 \), another piece of cortical tissue \( N_2 \) alters its organization in order to become type-identical to and possess the same representational properties as \( N_1 \) at time \( t_1 \). In the present case, tissue nativism is false and, by hypothesis, the type-identity thesis is true. But it may still be the cases that organism nativism is true. That is, it may be innately specified that the organism possess the representation \textit{red} because it may be innately

\[18\] Elman et al. sometimes appear to assume that mental representations are \textit{type identical} to neural entities: i.e. that every type of mental representation \( M \) is identical to a type of neural object \( N \). Thus they propose that we define the term ‘representation’ as ‘the fine-grained patterns of cortical activity, which in turn depend on specific patterns of synaptic connectivity’ (p. 364).
specified that some piece of cortical tissue—but not necessarily N₁ (or for that matter N₂)—encode red.¹⁹

6. Nativism and Neurobiology

So far I have argued that Elman et al. fail to undermine RN because they make a series of unwarranted assumptions about the commitments of the representational nativist. But perhaps they are to be forgiven for this misunderstanding. Nativists are typically far from clear in stating their views and work obviously remains to be done in order to clarify their position. With this in mind, I propose to spend the remainder of this paper discussing some prima facie plausible proposals for how one might understand nativism about representations and how—if at all—on these proposals, neurobiological evidence might bear on the question of whether or not RN is true. In particular, I’ll discuss two rather different approaches to trying to understand nativism and argue that on either of these, evidence from invasive, developmental, neurobiological studies is simply irrelevant to the question of whether or not RN is true. Thus, if we conceive of nativism in either of these ways, Elman et al. fail to undermine RN not merely because there are problems with the specific details of their argument but because the whole approach is doomed to failure: no argument that depends on invasive, developmental, neurobiological studies would warrant the conclusion that RN is false.

6.1 Neurobiology and Normal Environments

There are notorious problems with determining precisely how the notion of innateness ought to be explicated.²⁰ A wide range of proposed explications, however, invoke the concept of a standard (or normal) environment (Stich, 1975; Kitcher, 1996).²¹ What I want to argue is that on any account of innateness that invokes the notion of a standard environment, it will turn out that empirical data obtained from invasive, experimental studies on

---

¹⁹ Indeed, even if the representation did not ‘grow back’ it would not follow that organism nativism is false. Legs don’t grow back when you cut them off, but they are surely innate.

²⁰ The issue of how precisely ‘innate’ ought to be explicated cannot be addressed in the present paper. For reasonably sophisticated although incomplete discussions of what the notion of innateness means, see Stich, 1975; Block, 1981; Cosmides and Tooby, 1997; Griffiths, 1997, and Cowie, forthcoming.

²¹ For present purposes we do not require a precise explication of ‘standard environment’. The (very) rough idea, however, is that an environment E is normal (or standard) for a species S if it is relevantly similar to environments in which members of S typically develop. See Kitcher, 1996, pp. 246–8, for a more detailed discussion of the notion of a standard environment.
developing organisms are simply irrelevant to whether or not RN is true.\textsuperscript{22} In order to illustrate this point, consider the following, familiar characterization of innateness:

A characteristic is innate just in case it is ‘represented in and determined by the genes’ (Block, 1981, p. 280).

This characterization of ‘innate’ is of course vague and could be explicated in a variety of different ways. Specifically, the notion of genetic determination is in need of further clarification. One common characterization of genetic determination is that

A characteristic $C$ is genetically determined for an organism $O$ just in case organisms that have the same combination of alleles as $O$ develop $C$ in all standard environments.\textsuperscript{23}

Thus, genetic determination is characterized as a mapping relation between genes and phenotypic traits in some restricted range of environments—i.e. the set of standard environments.\textsuperscript{24}

With this characterization of innateness in hand, let’s consider the question of whether or not evidence from invasive, developmental, neurobiological studies could refute RN. The answer to this question, I maintain, is ‘No’. The reason becomes clear when we ask the question ‘Does an environment in which a neurobiologist performs invasive experiments on a developing organism count as a standard (or normal) environment for that (kind of) organism?’ So, for example, does an environment in which a neuroscientist opens up a human neonate’s skull and moves pieces of cortical tissue around (or rewires the afferent inputs to pieces of cortical tissue) count as standard? The answer is surely ‘No’. The environmental conditions under which humans have typically developed do not involve such invasive surgical procedures. Furthermore, such procedures alter species-typical patterns of development. According to the present conception of innateness, then, using

\textsuperscript{22} Of course, I do not intend to maintain that invasive, neurobiological experiments on developing organisms will tell us nothing whatsoever of interest, but only that on a standard environment analysis of innateness they won’t provide us with an answer to the question of whether or not RN is true.

\textsuperscript{23} This notion of genetic determination has recently been elaborated in Kitcher, 1996.

\textsuperscript{24} The reason for characterizing genetic determination in terms of standard environments as opposed to merely saying that $C$ is genetically determined for $O$ if it develops in all environments, is that it’s clear that no (sane) nativist would wish to claim that any human characteristic was genetically determined in this sense. After all, if innate, hence, genetically determined characteristics must develop in literally all environments, then presumably they should develop in an organism that is located, say, on the surface of the sun or when the developing embryo is exposed to a rich cocktail of drugs like thalidomide. But then it turns out to be obviously true that no human characteristics are genetically determined and, hence, obviously true that no human characteristics are innate.

© Blackwell Publishers Ltd. 1998
the results of such invasive experiments in order to argue against RN is rather like arguing against the claim that having only two hands and ten fingers is innate on the grounds that I would have developed no hands at all had I been exposed to thalidomide in utero. Environments in which I am subjected to levels of thalidomide sufficiently large to inhibit the growth of hands are not standard environments and nor are environments where my brain is rewired by a neurobiologist as part of a cortical plasticity study. Yet, if such environments don’t count as standard, then experimental data taken from invasive studies clearly won’t resolve the debate over whether or not a given characteristic is genetically determined and, hence, innate. After all, such experiments only tell us what happens in an environment which is, by hypothesis, non-standard.

To summarize: if we adopt the above characterization of innateness, then evidence from invasive, developmental studies is simply irrelevant to the question of whether or not RN is true. Moreover, the point can be generalized: according to any characterization of innateness that makes reference to standard (or normal) environments, evidence from invasive, developmental studies will be irrelevant to the question of whether or not RN is true. So, if we adopt an analysis of ‘innate’ that invokes the notion of a standard environment, then Elman et al. fail to undermine RN not merely because there are problems with the specific details of their argument but because the whole approach is doomed to failure.

No doubt the conclusion of the previous paragraph is one that the representational nativist would be happy to endorse. But it is worth noting that there are significant problems with the sort of standard environment analysis of innateness on which the argument depends. In particular, it’s hard to develop a standard environment analysis of ‘innate’ that does not lead to prima facie absurd consequences. Consider, for example, the following problem with the analysis of innateness outlined earlier in this section. Suppose we were to transplant a rat embryo into a human womb and discover (to our amazement) that it developed into a normally functioning human being. Presumably, the human womb is not a standard environment for a developing rat embryo. So, on our standard environment analysis of innateness, the fact that the rat embryo developed into a human being would be irrelevant to the issue of whether or not the rat’s traits are innately specified. But, on any intuitive understanding of what ‘innate’ means, this is surely absurd. Contrary to what the above analysis of ‘innate’ suggests, this is precisely the sort of evidence that would be relevant to whether or not the rat’s phenotypic traits are innately specified.25

25 Another problem with the standard environment analysis of ‘innate’ discussed in this section is that, according to the analysis and contrary to what is typically supposed, representations that are acquired by general-purpose learning processes could turn out to be innate. So, for example, on the analysis outlined above the concept water could turn out to be innate even if we employed a general-purpose learning mechanism in acquiring this representation, just so long as we acquired water in all the standard environments.

© Blackwell Publishers Ltd. 1998
6.2 An Issue-Based Approach to Understanding Nativism

At this point, there is a range of different options available to us in trying to understand nativism. One strategy might be to fine-tune the standard environment analysis discussed in section 6.1 so as to avoid the sort of problem I mention above. Another approach might be to develop an analysis of ‘innate’ that makes no use at all of the notion of a standard environment. For present purposes, however, I wish to put the issue of how to analyse the concept of innateness to one side and discuss an alternative and orthogonal approach to understanding nativism, which I call the issue-based approach. According to this approach, rather than trying to analyse ‘innate’ and its cognates, one instead attempts to understand nativism by identifying the main issues—the central bones of contention—that are in dispute between nativists and their opponents.

My current, best hunch is that there are a cluster of issues that are located in the vicinity of debates over RN.\footnote{Although I will not argue the point here, I suspect that on an issue-based approach RN resolves into several distinct questions of which the one I discuss in this section—viz. the size of the inventory of primitive representations—is merely one. Another issue that surely constitutes a central source of dispute between advocates of RN and their opponents concerns whether or not the mechanisms for the acquisition of representations are domain-general or domain-specific (Cowie, forthcoming).} For the moment, however, I will focus on only one issue which I take to be a central—indeed, perhaps the central—source of dispute between representational nativists and their opponents in contemporary cognitive science. Roughly put, the issue is this: are there any, and if so how many, representations that need to be posited as primitives of our psychological theories? Nativists invariably claim that there are a lot of representational primitives, while non-nativists maintain that there are few if any such primitives.

A few points of clarification and elaboration are in order. First, what do I mean when saying that a representation is a primitive of our psychological theories? While the term ‘primitive’ is used in a variety of different ways in the literature on nativism, for present purposes I intend it to be understood in the following manner:

A representation R is a psychological primitive if:

1. Some correct psychological theory posits the existence of R and
2. No intentionally characterizable mechanism or process accounts for the acquisition of R—i.e. there is no (correct) intentional psychological theory of how we acquire R.\footnote{There is, of course, an interesting issue concerning precisely when a process or mechanism is a genuinely intentional one. But for the moment I’ll put this issue to one side.}

So, for instance, according to many theories of concept acquisition, most

\copyright Blackwell Publishers Ltd. 1998
concepts are acquired via a process of hypothesis formation and testing (Bruner et al., 1956; Fodor, 1981). Since hypothesis formation and testing is a paradigmatic example of an intentional process, concepts that are acquired by this method are not psychological primitives in the above sense. Concepts that are acquired, however, via a process of triggering will be psychologically primitive, since one central characteristic of triggering is that it is supposed to be a non-intentional process.

Second, why think that the dispute over the size of our inventory of psychologically primitive representations is a central point of dispute between representational nativists and their opponents? There is a long answer to this question that involves a detailed analysis of the theories and explanatory practices of nativists and their opponents. But the short story is that many central figures in debates over RN often maintain that innately specified representations are acquired by ‘brute-causal’ and, hence, non-intentional mechanisms as opposed to ‘rational’, hence, intentional, psychological mechanisms (Chomsky, 1980; Fodor, 1981). Thus, for example, Fodor maintains that: ‘A Nativist says that the mechanism underlying the acquisition of all lexical concepts is brute-causal’ whereas

The empiricist model says that ... there are, as one might say, rational-causal processes: the psychological mechanisms which mediate the availability of such concepts realize an inductive logic. (Fodor, 1981, p. 275)

The claim that mechanisms for the acquisition of concepts or representations are ‘brute-causal’ is clearly intended to capture the idea that there is no correct, intentional psychological story of how such mechanisms operate—that our psychology will be silent on the matter of how we came to possess the concept. In contrast, inductive reasoning processes are a paradigmatic instance of intentional psychological ones. So, according to Fodor, the dispute over the size of the inventory of psychologically primitive concepts is at the heart of the debate over concept nativism.

The final issue we need to address concerns the relationship between this issue about the size of our inventory of psychologically primitive representations and the kinds of neurobiological evidence that Elman et al. invoke in order to undermine RN. Specifically, does the neurobiological evidence provide us with a way of settling the dispute between nativists and their opponents over how large our inventory of psychologically primitive representations is? Once again, I maintain that the answer is ‘No’. And this is

28 Though I will not argue for the point here, it should be reasonably obvious that disputes between representational nativists and their opponents in cognitive science frequently centre on issues about the size of our inventory of psychologically primitive representations. So, for example, the dispute between Chomskians in linguistics and their opponents often takes the form of a dispute over whether or not one needs to posit a set of primitive representations that encode for grammatical knowledge.
simply because the kinds of evidence invoked by the neurobiological argument are silent on the matter of whether or not the mechanisms (or processes) responsible for the acquisition of representations are specifiable in intentional terms. What the neurobiological argument does show is that experimental manipulations can alter the representational properties of pieces of cortical tissue. But this clearly does not tell us whether or not the mechanisms responsible for acquiring representations can be intentionally characterized. So, on the present approach to understanding RN, the neurobiological argument won’t settle the question of whether or not RN is true.

7. Conclusion

I have argued that the neurobiological argument fails to undermine RN because it makes a series of unwarranted assumptions about nativism and about the extent to which neurobiological data constrain claims about the innateness of mental representations. Pace Elman et al., the innately specified properties of a piece of cortical tissue need not be intrinsic properties of the tissue, nor must they satisfy the Principle of Invariance. Moreover, even if we were to assume that the innate properties of neural tissue must satisfy both of these conditions, the neurobiological argument would still fail to undermine the version of RN that psychologists actually defend—it would only provide us with reason to reject tissue nativism and not organism nativism. Finally, if we adopt any of a range of characterizations of ‘innateness’ that utilize the notion of a standard environment or else construe nativism as a dispute over the size of our inventory of psychologically primitive representations, then no argument that employs evidence from invasive, developmental, neurobiological studies will refute RN because all such evidence would be irrelevant to the question of whether or not RN is true. Contrary to the claims of Elman and his co-authors, then, the neurobiological data provide us with absolutely no reason to conclude that RN is false. Indeed, if what I have said is correct, it is very hard to see how data from developmental, neurobiological studies could refute RN. Though the recent progress of neuroscience has been truly impressive, neurobiology is not going to settle the debate over innate representations.

Department of Philosophy
University of Pennsylvania

References


© Blackwell Publishers Ltd. 1998