

What Does It Mean to Claim that Something Is ‘Innate’? Response to Clark, Harris, Lightfoot and Samuels

ANNETTE KARMILOFF-SMITH, KIM PLUNKETT,
MARK H. JOHNSON, JEFF L. ELMAN AND ELIZABETH A.
BATES

From the staunchest Chomskyan nativist to the most domain-general empiricist, all scientists concur that development involves contributions from both genes and environment. The ravine between theories lies in how rich and domain-specific the innately specified component is claimed to be, how directly genes or environment contribute to developmental outcomes, and whether the environment acts merely as a trigger to pre-specified knowledge. When theorists claim, for instance, that the child is born with the knowledge that human languages contain nouns and verbs (e.g. Pinker, 1994), what do they mean precisely? Is this simply a loose metaphor, or are they seriously claiming that such knowledge is pre-encoded in genes that map directly to cortical circuits? *Rethinking Innateness* (henceforth *RI*) attempted to clarify the whole issue of what it really entails to claim that something is innate.

While Clark's review of *RI* is generally very positive, he raises two essential points. First, he argues that our distinction between representational innateness (e.g. pre-specified knowledge content) and architectural innateness (e.g. pre-specified learning mechanisms) should lie on a continuum and not be seen as a dichotomy. Second, he submits that we do not have evidence to support our claim that representational innateness must be synonymous with pre-wiring of synaptic connectivity. We will reply to the second issue first, since it underpins our response to the first. If we take the term 'innate knowledge' seriously (e.g. knowledge, say, of the abstract constraints that obtain on human languages being available to the child prior to linguistic experience), then these must be encoded in the brain as a particular pattern of synaptic connectivity within a specific neural system. This

is not only our understanding of 'knowledge' but that of others such as Pinker (1994). However, neurobiological data indicate that cortical micro-circuitry develops in an experience-dependent way (e.g. Shatz, 1992, 1994), despite the fact that there are endogenous (non-activity-dependent) constraints on connections at the level of gross input-output wiring (axonal connections).

This is not necessarily a claim about the specific location of neural tissue, a misunderstanding that permeates Samuels' contribution. While philosophers of a nativist persuasion might go along with 'tissue vs. organism' nativism, most cognitive neuroscientists would not. Cerebral cortex is critical for higher cognitive functions. While agreeing that *RI* helps to reject tissue nativism, Samuels replaces this with the notion of 'organism nativism'. But this latter concept is simply too underspecified to be of any use scientifically, even if it lends itself to philosophical musings. Samuels cites a common characterization of genetic determination:

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.

But simply noting that X (e.g. language) occurs in humans and not in other species does not lead to the conclusion that X is innate. To say anything scientifically useful, one must be explicit about the pathway from genome to phenotypic outcome.

Samuels also needs to define what he means by 'standard environment'. In fact, this general notion has already been discussed in great detail by non-nativists in terms of experience-expectant processes (Greenough, Black and Wallace, 1993) or species-typical environments (Johnson and Morton, 1991). Samuels argues that any change to the standard environment renders a study invalid in terms of refuting nativism, since his characterization of nativism includes the so-called standard environment. This hinges on a very slippery notion of environment.

In any case, we (and many others) argue that neurobiological constraints are in fact frequently ignored and yet they should be taken into account when theorists define what they mean by 'innate'. So-called 'tissue nativism' is extremely common in scientific journals and textbooks. In fact, the study of specific lesions in brain-damaged adults regularly uses neural specificity in support of the nativist position. The major aim of nativist-inclined developmentalists is to map cortical areas and behavioural/cognitive functions and, using the method of double dissociation, to make nativist claims about the starting state and the role of genetic coding. *RI* by no means invented a straw man in the form of tissue nativism. Indeed, as Samuels himself points out in his introduction, the doctrine of innate ideas or 'representational nativism has become an important theoretical option in cognitive science and has been invoked in order to explain a wide range of phenomena'. Representational nativism is alive and well. But it ignores development and

leads to a questionable evolutionary hunt for de-novo domain-specific traits rather than adaptations of existing ones. Having a vocal tract that makes a right-angle bend is very useful for language because it enriches the repertoire of sounds that can be produced. But did evolution create this angle 'for language'? Probably not, because it is also the result of upright bipedalism. However, it does have a domain-specific *consequence*. We should not equate starting state and outcome when considering either evolution or ontogeny. As developmental scientists, we seek not to describe the endstate, but to understand *how* the higher cognitive functions in neocortex progressively come to be what they eventually are.

In *RI* we attempted *inter alia* three things: (a) to determine the space of *logical* possibilities for constraining knowledge, (b) to ascertain which of those possibilities are plausible, given what is known about the neurobiology of development, and (c) to understand the precise implications of the nativist claims abundant in the literature. Rather than using the term 'innate' in an unspecified way biologically, we broke it down into representational, architectural and chronotopic constraints and assessed the literature in this more precise light. We started with the premise that mental activity (the cognitive level) must be equated with brain activity, and that the primary determinant of brain activity is the pattern of inter-neural synapses. (To simplify things in *RI*, we did not consider many other important factors which influence neuronal activity, such as hormonal and diffuse neuromodulatory substances.) If knowledge does not reside in a set of synaptic weights, how, we ask, could such specific knowledge be encoded in the brain prior to experience? It should be noted that in *RI* we do *not* claim that the only way to constrain knowledge is to build specific patterns of cortical synapses in advance of activity. What we do recognize is that this is nonetheless a logical possibility and that there are (albeit rarely) species for which such patterns of connectivity obtain (e.g. *C. Elegans*). But we stress in *RI* that many theorists do make such claims about pre-specified patterns of connectivity with respect to the human species—see, for instance, the numerous nativist citations in chapter 7 of *RI*. One positive outcome of *RI* would be to encourage others to deny that they ever believed in representational nativism, but then to force them to define precisely what they mean by 'innate knowledge'. We would welcome the retreat of theorists, particularly in the area of Chomskyan linguistics, from an extreme nativist position into embracing one of more 'sensible nativism', to use Clark's terms. But Clark is wrong in refusing to acknowledge that representational innateness is both a logical possibility and one which currently still has a large number of very committed adherents.

In our view, the issue cannot be surmounted simply by blurring the distinction between representational and architectural innateness. We believe that one must distinguish between pre-wired patterns of synaptic connectivity and patterns of synaptic connectivity which arise as a result of developmental biases. We argue that development itself is the key to understanding how representations emerge. A connectionist architecture with a random set

of weights has no prior 'knowledge' at all, initially knowing nothing about the mapping function (the knowledge) that it will end up performing. Its knowledge accrues as a result of progressively processing the input. Of course, knowledge can be built in by setting weights in advance, but it has been repeatedly shown that without gradual learning, the kinds of interesting errors typical of real children will not occur in network simulations.

Clark argues that, despite the fact that architectural and timing biases which affect, say, language may also have effects on other acquired skills, it is still appropriate to speak of such biases as being language-specific since they might have been subject to selection precisely because they favour that domain. A useful distinction in this respect is between domain relevance and domain specificity (Karmiloff-Smith, 1998). A system can have domain-relevant architecture, i.e. the architecture will be more suitable to processing a certain type of input than others (say, a recurrent network for processing sequential input, Elman, 1993), although the architecture is capable of processing non-sequential input. However, given competing inputs and competing architectural constraints, domain relevance may force the network into focusing more on a certain type of input. The network will thus *become* domain-specific as a result of progressively processing the structure of that input. In other words, domain-relevant architectural constraints will have domain-specific consequences. But this doesn't mean that the knowledge has to be built in in advance; the domain-specific knowledge will be acquired as a result of the network's learning history. Pinpointing differences in firing thresholds, types of neurotransmitter, packing density of neurones, degree of macro-connectivity—any of which might make some brain circuits more suitable for language learning than others—is quite different from claiming genetic encoding of knowledge of Universal Grammar in the micro-circuitry of the brain.

Clark's suggestion to blur the distinction between representational and architectural nativism thus fails to capture the fundamental (philosophical and scientific) distinction between knowing something a priori and having the tools with which to discover something. So although Clark seeks a middle ground by suggesting that no useful purpose is served by our quibbling over the distinction between domain-relevant biases and domain-specific representational outcomes, we believe the distinction to be extremely important.

That being said, Clark does present the authors of *RI* with a possible challenge because there does exist a scenario in which the distinction between representational and architectural nativism is blurred. Let's suppose a neural system starts life with a set of random weights. Further suppose that the general pattern of connectivity and learning algorithms are such that neural activity drives weight adaptation to a target configuration *irrespective of the specific activity patterns experienced*. In such a case, the difference between representational and architectural innateness becomes blurred, because the environment is merely a trigger for pre-specified weight adjustments. However, no cortical system at the cognitive level has yet been shown to function

in this way. Indeed, current neurobiological research points to the contrary. The discussion can go no further until Clark produces a concrete example regarding any cognitive system that develops irrespective of the activity patterns it experiences. Of course, this position is held by staunch nativists like Lightfoot and Samuels who claim that the environment acts 'only by triggering the representation, rather than by playing the role of evidence in a learning process'.

Harris focuses on a more specific issue: vocabulary development. She is concerned that the emergent properties observed in the vocabulary learning network are artefacts of the training regime—a general criticism of connectionist commentators in the past. She rightly points to the fact that, at least in the early stages of vocabulary development, children are unlikely to learn label-image pairs simultaneously. They will know quite a lot about the world before attaching phonological forms to objects, and conversely they may learn a phonological form and take some time to work out what it applies to. Harris's concern is justified. The model in question differs from real learning in that it is trained on a large number of simple image-label pairings in each training epoch. What, for instance, would happen if the model were trained in an incremental fashion, with few image-label pairs initially and then expanded to incorporate all object-label pairings? Would the model still exhibit the same emergent properties (vocabulary spurt, comprehension/production asymmetry, over- and under-extension errors, prototype effect)? Although such a simulation has not been run, we already know enough about the general properties of connectionist systems and the specific results of incremental learning (Elman, 1993; Plunkett and Marchman, 1993) to predict what would happen in the case of vocabulary development. Starting out with a small set of image-label pairs will result in strong learning of those pairs, with prototype effects emerging as the range of images associated with a given label is extended. The comprehension/production asymmetry will be unaffected by the incremental training regime because it is a by-product of the many-to-one mapping between objects and labels and the greater sparseness of label representations as compared to object representations. The vocabulary spurt will be delayed compared to mass training, but the same shape of change will be evident (cf. chapter 4 of *RI*, as well as Plunkett and Marchman, 1993). Finally, over- and under-extension errors would still occur, with under-extension errors most evident prior to the vocabulary spurt and over-extension errors after the spurt. Age-of-acquisition effects would be more, not less, apparent with incremental training, with early image-label pairings showing more robust learning and fewer errors. So, it turns out that Harris's concerns that a different training regime would not give rise to the same emergent properties lead to the contrary conclusion: an incremental training regime would give rise to a better approximation to the kinds of emergent properties seen in real children.

A second issue raised by Harris concerns individual differences. Can these be accommodated within a connectionist framework? *RI* neglected to

address this important issue in any depth, but it has since been dealt with in some detail (Plunkett and Marchman, 1996; Plunkett, Karmiloff-Smith, Bates, Elman and Johnson, 1997). If connectionist models (e.g. of past-tense acquisition) are trained on different rates of vocabulary increment, they display variations both in the onset of over-regularization errors and in the rate of acquisition of regular and irregular verbs. In general, connectionist models can vary due to both exogenous factors (the training environment) or endogenous factors (different learning rates, different numbers of hidden units, and so forth), all of which can alter developmental trajectories. Likewise, in real children, amount and quality of input may differ as may each child's endogenous capacity to learn. Quantitative rather than qualitative variations in architecture and timing may explain individual differences. Altering different parameters in a connectionist network will also produce individual differences in onset of vocabulary spurt or its abruptness. But doing so in networks helps us to pinpoint the mechanisms responsible for such differences. As Harris rightly states, 'accounting for developmental change, rather than merely describing it, is the major task for developmental psychology, and the adequacy of any theory of child development must be judged in terms of the adequacy of its account of change'. *RI* illustrates how the precise constraints of connectionist modelling can serve to explore in detail the mechanisms by which individual differences occur in real children.

Finally, while being enthusiastic about what *RI* attempts, Harris expresses a more general concern about psychological reality. This of course holds for all models, be they nativist, box-and-arrow information processing, or connectionist. Any model must be evaluated on the basis of whether it clearly states its assumptions about the fundamental principles at work in the domain of interest, whether it approximates real data and, perhaps most importantly, whether it generates novel empirical predictions. Throughout the chapters of *RI*, we confront all three of these theoretical constraints.

With respect to Lightfoot's commentary, we are flattered that as a reaction to *RI* he went to such extended and convoluted ends to attempt to save an otherwise doomed philosophical position. We also assure him that there was nothing 'uncomfortable' about our collaboration. He wrongly assumed that *RI* was a juxtaposition of parts written by different authors. It was a collaborative effort throughout.

Lightfoot appears to be totally unfamiliar with the burgeoning field of computational neurobiology and the large number of journals publishing works on neural networks modelled on the properties of single neurones or groups of neurones. He seems to believe that neuroscience is still rooted in its Aristotelian days! He also makes the typical error of equating constructivism with empiricism and confusing triggering with interaction. Indeed, to support his staunch nativist claims, Lightfoot offers the reader several contradictory statements, amongst which: (a) nobody ever doubted that children learn in any normal sense of that term, and (b) that Piattelli-Palmerini's 1980 discussion of innateness is far more substantial than the discussion in *RI*.

Yet it is precisely Piattelli-Palmarini who claimed (1989) that there is no such thing as learning!

Lightfoot further claims that: (a) nativist linguists have long embraced an interactionist position, and (b) this interaction consists of 'genetic contributions of Universal Grammar, represented generally in the species, and the specific triggering experience provided by the linguistic environment'. The notion of interaction normally implies that *both* sides actually contribute to the structure of the developmental outcome—not that one side merely triggers the other's pre-specified constraints. Gottlieb (1992) has made this distinction very clear: if the environment merely acts as a trigger, then we are dealing with predetermined epigenesis, but for both genes and environment to play a truly interactive role, we must think in terms of probabilistic epigenesis. It is perhaps clearer to specify that our notion of interaction is not in fact between genes and environment (Karmiloff-Smith, 1998). Rather, on the gene side, the interaction lies in the outcome of the indirect, cascading effects of interacting genes and their environments and, on the environment (experience) side, the interaction comes from the infant's *progressive* selection and processing of different kinds of input. For both the strict nativist and the empiricist, the notion of 'environment' is a static one, whereas development (both normal and atypical) is of course dynamic.

Like Samuels, Lightfoot constructs such a broad notion of nativism that it is indeed almost impossible to refute. It is, however, also exceedingly difficult to prove. Whether such a broad concept is philosophically useful remains questionable, but its scientific utility is totally unclear. Throughout his paper, Samuels has the term 'innate' do a lot of work for his argument, yet he explicitly refuses to define it! Surely even in the field of philosophy—or perhaps especially in a field such as philosophy—this is not fair play. Both authors ignore the message hammered home throughout *RI* that nativists must be more explicit about what they mean by innate and what the implications of their nativist positions are biologically. Does Lightfoot really believe that Universal Grammar is encoded directly in the genes? Is there any evidence for this, other than philosophical contemplation? And stating that something is 'in the genes' is simply not sufficient if one is not simultaneously explicit about how gene expression takes place. By contrast, in *RI* we made a commitment to defining what we meant by 'innate' and what we took others to mean, and we showed how, if used at all, the term must distinguish between representational, architectural and chronotopic innateness. One can disagree with our definitions, but alternative explicit definitions must be provided if we are ever to take the debate scientifically beyond its current level.

Lightfoot rightly points out that 'there is nothing in the architecture of connectionist models which would deny the possibility of domain-specific devices'. Of course. Connectionism is a tool, not directly a theory, and, as pointed out above, staunch nativist views can be explored within a connectionist framework by building in weight connections from the outset. It is most odd that Lightfoot dubs the connectionist past-tense model as 'domain-

specific'. Is he confounding the emergent end point of learning with the starting state? Has he completely missed that distinction? Surely it is obvious that the model used for the past-tense simulation could have learned to balance beams or catch balls or transform active sentences into passive sentences. Its architectural and computational constraints would have allowed this. The point is that it is the structure of the input that also plays a vital role in the outcome of learning. If weights had been pre-specified, then indeed it would only have been able to learn about one particular domain. But the model was a domain-general device being put to work on a domain-specific task. The whole point of *RI* was to show that some kinds of architectural constraints enable the learning of some kinds of input, and other kinds of architectural constraints are more suitable for different kinds of input. But these are many-to-many mappings, not the one-to-one mappings that Lightfoot, Samuels and like-minded theorists seek. Surely Lightfoot realizes that the field has moved on dramatically since the initial Pinker and Prince criticism (1988) of the Rumelhart and McClelland (1986) model of the past tense a decade ago! But even if the field had remained static, a single-layered perceptron is not a domain-specific learning device, so even here Lightfoot's criticism is misplaced.

Finally, in stating that 3–5-year-olds are 'at the earliest stage that can be tested' in language, Lightfoot shows his ignorance of the enormous advances in infancy research. It has now been clearly demonstrated that from the beginning of life infants display an enormous capacity for learning, including statistical learning of sequential patterns after only a few minutes of exposure (for visual learning in 3-month-olds, see Haith, Hazan and Goodman, 1988; for auditory learning in 8-month-olds, see Saffran, Aslin and Newport, 1996). As Elman (1990) has shown, any organism capable of prediction can generate its own negative evidence; hence the demonstration that infants can predict the elements within a sequenced event constitutes a powerful challenge to one of the strongest nativist arguments against the possibility of learning in complex domains (including grammar). The average 3-month-old infant has had approximately 900 waking hours or 54,000 minutes of auditory and visual experience, more than enough to support substantial learning about the world. And these calculations do not even take into account the last three months of intra-uterine life, when auditory learning is already seriously underway (DeCasper and Fifer, 1980). Against this background, Lightfoot's belief that the existence of grammar in 3-year-olds constitutes a proof of innateness is, quite simply, silly. Relevant to all this is another evolutionary fact that staunch nativists need to take into account: the structure of the human infant's brain continues to develop over many years, allowing the environment to play a very active interactional role. This applies particularly to the micro-circuitry of neocortex where higher cognitive functions operate (but *not* to the infant's head and body, which change in size only, so the analogies drawn by Samuels are invalid in this respect). In short, many of the classical arguments of linguists and

philosophers of a nativist bent simply fail in the face of all the new empirical facts about infancy.

In conclusion, we believe that it is high time that our field went beyond the polarized debates about innateness and domain specificity. Neurobiology provides us with a consistent message: organisms develop in activity-dependent ways. This is why questions concerning innateness must be asked within a truly developmental framework and must be underpinned by *explicit* definitions of what it means to claim that something is 'innate'. In our view, *RI* has provided a sensible framework within which to further such discussion.

*Neurocognitive Development Unit
Institute of Child Health, London*

*Department of Experimental Psychology
Oxford University*

*Psychology Department
Birkbeck College, London*

*Centre for Language Studies
University of California at San Diego*

References

- DeCasper, A. and Fifer, W. 1980: Of Human Bonding: Newborns Prefer Their Mother's Voices. *Science*, 208, 1174–6.
- Elman, J. 1990: Finding Structure in Time. *Cognitive Science*, 14, 179–211.
- Elman, J. L. 1993: Learning and Development in Neural Networks: The Importance of Starting Small. *Cognition*, 48, 71–99.
- Gottlieb, G. 1992: *Individual Development and Evolution: The Genesis of Novel Behaviour*. Oxford University Press.
- Greenough, W. T., Black, J. E. and Wallace, C. S. 1993: Experience and Brain Development. In M. Johnson (ed.), *Brain Development and Cognition: A Reader*. Oxford: Blackwell, 290–322.
- Haith, M., Gazan, C. and Goodman, G. 1988: Expectation and Anticipation of Dynamic Visual Events by 3.5-Month-Old Babies. *Child Development*, 59, 467–79.
- Johnson, M. H. and Morton, J. 1991: *Biology and Cognitive Development: The Case of Face Recognition*. Oxford: Blackwell.
- Karmiloff-Smith, A. (1998). Development Itself Is the Key to Understanding Developmental Disorders. *Trends in Cognitive Sciences*, vol. 2, no. 10, 389–98.
- Piatelli-Palmarini, M. 1989: Evolution, Selection and Cognition: From 'Learning' to Parameter Setting in Biology and the Study of Language. *Cognition*, 31, 1–44.
- Pinker, S. 1994: *The Language Instinct: How the Mind Creates Language*. New York: William Morrow.

- Pinker, S. and Prince, A. 1988: On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition. *Cognition*, 28, 73–193.
- Plunkett, K. and Marchman, V. 1993: From Rote Learning to System Building: Acquiring Verb Morphology in Children and Connectionist Nets. *Cognition*, 48, 21–69.
- Plunkett, K. and Marchman, V. 1996: Learning from a Connectionist Model of the Acquisition of the Past Tense. *Cognition*, 61, 299–308.
- Plunkett, K., Karmiloff-Smith, A., Bates, E., Elman, J. L. and Johnson, M. H. 1997: Connectionism and Developmental Psychology. *Journal of Child Psychology and Psychiatry*, 38, 53–80.
- Rumelhart, D. E. and McClelland, J. L. 1986: On Learning the Past Tense of English Verbs. In D. E. Rumelhart and J. L. McClelland (eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* vol. 1, *Foundations*. Cambridge, MA: MIT Press, 318–62.
- Saffran, E. M., Aslin, R. N. and Newport, E. L. 1996: Statistical Learning by 8-Month-Old Infants. *Science*, 274, 1926–8.
- Shatz, C. J. 1992: Dividing Up the Cortex. *Science*, 258, 237–8.
- Shatz, C. J. 1994: How Are Specific Connections Formed between Thalamus and Cortex? *Current Opinion in Neurobiology*, 2, 78–82.