Chapter 8

Innateness and the Situated Mind

Robert Rupert

Many advocates of situated approaches to the study of cognition (e.g., Griffiths and Stotz, 2000; Thelen and Smith, 1994) explicitly take exception to cognitive science's pronounced nativist turn.[1] Other proponents of situated models seek to mitigate strong nativist claims, by, for example, finding ways to acknowledge innate contributions to cognitive processing while at the same time downplaying those contributions (Wilson, 2004, Chapter 3). Still others leave implicit their apparent opposition to nativism: they emphasize the environment's contribution to cognition so strongly as to suggest antinativist views but do not take up the issue explicitly (Clark, 1997; Varela, Thompson, and Rosch, 1991).[2] Thus, situated theorists have reached something approximating an antinativist consensus. In this chapter, I argue that they should not embrace the antinativist view so readily. To this end, I divide the situated approach into two species, extended and embedded views of cognition, arguing that each version of the situated view admits of a plausible nativist interpretation with respect to at least some important cognitive phenomena. In contrast, I also argue for the nonnativist interpretation of certain cognitive phenomena; nevertheless, these antinativist recommendations come heavily hedged -- in some cases, at the expense of a robust reading of the situated program or one of its subdivisions.

I. Extended Cognition and Nativism

Consider first the view that cognitive processes extend beyond the boundary of the organism. The intimacy of the human organism's interaction with its environment during cognitive processing suggests that those cognitive processes literally comprises elements of the environment beyond the boundary of the human organism (Clark and Chalmers, 1998). I shall refer to this view as the 'hypothesis of extended cognition', or 'HEC'. As I understand it, HEC entails that the human *mind* is extended. Accordingly, the subject matter of HEC -- i.e., the kind of cognition at issue -- had better be the sort that bears on the location of the mind. This seems fair enough. The explananda of cognitive science are various mental capacities broadly to do with belief formation, such as the capacities to reason, perceive, remember, construct theories, and use language. Thus, whatever model of cognition we ultimately adopt will be a model of the mind's activities or capacities; and if the activities of a mind take place at a particular spatio-temporal location, that mind is at least partly at that location.[3] On this view, a given mind has a location in space-time and, according to HEC, this location includes points outside the 'skin-bag', as Andy Clark (2003) has colorfully dubbed the boundary of the human organism. Thus we arrive at the division mentioned above. A view is extended if it holds that in some cases, a system composed of a human organism together with material (possibly including other organisms) existing beyond that organism's boundaries instantiates cognitive properties relevant to the location of a mind; views that merely emphasize the human organism's heavy dependence on, and frequent interaction with, the environment during cognitive processing are embedded views, discussion of which is deferred to later sections.

If extended systems are the proper objects of study in cognitive science, antinativism seems to follow, particularly if one accepts the common notion that a trait is innate if and only if it is specified by the genome (Block, 1981, pp. 280-81; Elman et al., 1996, p. 22). The advocate of extended cognition urges us to focus on the traits of extended systems, and it is difficult to see how genes could encode such traits, for genes would seem directly to affect only the organism itself. Extended theorists typically hold that cognitive systems include such external components as hard-drives, notebooks, text messages, and, in the case of vision, whatever the subject happens to be visually engaged with. How, one might wonder, could genes specify anything about such external resources? These extended elements are disconnected from the organic milieu of the genome and thus beyond the genome's direct causal or informational purview. Given the great variability in the external resources alleged to become part of extended cognitive systems, it would seem that not even the entire set of biological resources internal to the organism can specify what will become part of the resulting extended cognitive system.

This antinativist argument from HEC seems quite powerful. What, in contrast, might motivate a nativist reading of HEC? Consider two widely discussed measures of a trait's or capacity's innate status: canalization (Ariew, 1999) and generative entrenchment (Wimsatt, 1999). The canalization-based approach to nativism holds that a trait is innate to the extent that it resists perturbation across changes in the environment. As an account of innateness, this allows a trait to be more or less innate, depending on the breadth of environments in which the trait appears.

Now consider some systems typically claimed to involve extended cognition: a human together with a map (Hutchins, 1995); a human together with a pencil and paper on which she performs mathematical calculations (Clark, 1997); a human together with auditory patterns of spoken language (Clark, 1997; Rowlands, 1999); a human together with the visible objects in her immediate environment (O'Regan, 1992); a human together with gross physical structures in her environment, for instance, a human and a roadway she is following (Haugeland, 1995).

At least some traits of such systems exhibit a high degree of canalization. Take, for example, a human using external symbols – say, self-directed speech – to guide herself through a complicated task (Clark, 1997, 1998). Humans typically live in groups and use language, and thus the typical human subject is likely to engage in some kind of self-directed symbol use across a wide range of environments. Furthermore, even in the absence of a clear channel of learning from conspecifics, humans tend to create linguistic or quasi-linguistic symbol systems (Bloom, 2000, chapter 10; Goldin-Meadow and Zheng, 1998). To the extent that subjects developing in abnormal environments use partly externalized, self-devised systems to guide their own thought and behavior, the self-directed use of a symbolic system passes the deprivation test for canalization (Ariew, 1999): deprived of a species-typical environment, the trait emerges nonetheless. The use of the environment as visual memory is, perhaps, even more broadly canalized; this trait appears in all but the most extreme environments – e.g., those involving physical damage to the visual system or restriction to dark environments. Once we have placed extended cognitive systems on par with organismic systems, the canalization theory of innateness delivers a nativist vision of many of the former systems' traits.[4]

Move now to William Wimsatt's theory of generative entrenchment (Wimsatt, 1999).[5] According to this view, a cognitive capacity or trait is innate to the extent that its appearance is a prerequisite for the appearance of traits emerging later in development. Some important and *prima facie* cognitive characteristics of extended systems seem to satisfy the criterion of generative entrenchment. Consider, for example, the human organism's spoken out, rote learning of times tables. This is a prerequisite for the acquisition of many later-emerging traits of the extended cognitive system. Typically, when learning the times tables, subjects intentionally create external sound (or print) structures to facilitate learning. For many subjects, either the continued vocalization of basic facts about multiplication or the internal representation of such verbalization proves invaluable in the solving of more complex problems (compare the intentionally created, linguistic means of external control discussed by Clark – 1997, pp. 195-96, 1998, pp. 173, 181). At the very least, one would think that, for any given subject, there must have been a time when the extended trait appeared; it is difficult to learn the times tables without verbalizing them (or writing them out – another extended process). The ability to solve problems involving the multiplication of large numbers depends on the memorization of the times-tables, and the ability to solve problems involving the multiplication of large numbers grounds many further abilities, for example, those underlying feats of engineering and commerce (cf. Wimsatt's discussion of generative entrenchment and cultural evolution, 1999, p. 143).

The antinativist advocate of HEC might respond directly, attempting to show that when properly applied, considerations of canalization and generative entrenchment do not support a nativist reading of HEC's favored examples. It is, however, difficult to execute this strategy without relegating external resources to second-class status, crossing purposes with HEC; for the most promising antinativist response invokes asymmetries in the contribution of the organism to the apparent canalization and generative entrenchment. It is not a trait of the extended system that is, e.g., canalized, so much as the ability of the human organism to enter into certain relations with external resources (cf. Clark, 2003); the appearance of alleged extended systems depends on the latter ability. Furthermore, this dependence is asymmetrical: the intentions, desires, and purposes of the organism provide impetus missing from external resources (I discuss this point in more detail below). This asymmetry implies that the organismic portions of extended systems are deeply privileged, to an extent that should unsettle the advocate of HEC (cf. Rupert, 2004). The antinativist HEC-theorist might do better, then, to advocate for a distinct biological perspective on innateness. Of most promise would seem to be developmental systems theory, or DST, which appears not only to preserve HEC's antinativist credentials but also to support directly the HEC-style individuation of cognitive systems.

Developmental systems theory is often seen as an antidote to twentieth-century biology's over-emphasis on the gene. For many years, textbook presentations characterized genes as the codes for and the determinants of phenotypic traits of living things; these encodings were selected for because they determined the presence of phenotype traits conferring reproductive advantage on their bearers. In contrast, the advocates of DST point to the wide range of contextual factors affecting gene expression and, more generally, the development of phenotypic traits. The most uncontroversial of these contextual factors reside in the organism itself. Developmental systems theorists go one further, however, arguing that determination involves the entire host of factors shaping the phenotype – i.e., the entire host of factors that create a life-cycle – which, in many cases, includes factors in the environment beyond the organism's boundary (Griffiths and Gray, 2004).

Developmental systems theory emphasizes the ways in which organismic resources contribute to the shaping of environments (in niche construction, for example) that confer reproductive advantage on the very organismic resources that help to create those environments. As a result, those environments are more likely to be recreated or maintained. Thus, environmental factors can themselves exhibit the dynamics of selection as it is thought to operate on genetic resources: the environmental resources exhibits traits that, given the context, increases their own fitness. In fact, the environments are selected for along with the selection for organisms that create those environments. Because reproduction of the two systems – of organismic resources and external ones – rely on each other for success, selection is for the composite package of resources, internal and external: the entire developmental system, organism-and-environment, is selected for.[6]

Developmental systems theory thus appears to support HEC in a fairly straightforward way. According to DST, evolutionary forces often operate on transorganismic, or extended, systems. Assuming that cognitive traits were selected for, it is no surprise that such traits should be exhibited by transorganismic systems. The HEC theorist simply takes the systems instantiating cognitive properties to be, or to at least be similar in scope to, the systems of fundamental importance in respect of the biological processes that give rise to cognitive phenomena.

What is more, DST seems to ground a non-nativist account of the cognitive traits of extended systems. Traditional notions of innateness presuppose an important distinction between the organism (or the mind) and its environment; nativists claim that, in some substantive sense, internal structures (beliefs, theories, concepts) arise independently of the environment (or of what is external to the mind) and are innate for just that reason. According to DST, evolutionary processes select for traits determined by more than the organism itself. On the traditional view of innateness, then, where there is extra-organismic determination of phenotypic traits, the traits in question are not innate. Similarly, in the case of cognitive systems: if a cognitive system includes the environment in nontrivial ways – ways that pertain directly to the cognitive capacities or structures in question – this precludes a nativist account of those structures; for surely the cognitive capacities in question are not sufficiently independent of the environment to be innate.[7]

Both appeals to DST fail the HEC-theorist. Take first the non-nativist interpretation of DST and, derived from it, the non-nativist interpretation of HEC. If there is any single, overarching idea driving nativist thought throughout the centuries, it is the equation of innateness with what follows from the *internal* properties or states of a system. If, however, the relevant system is extended, the internal-external boundary shifts accordingly: it separates what is internal to that extended system, including its organismic and extra-organismic parts, from what falls outside that extended system. It follows from HEC, then, that traits arising exclusively or largely from what is internal to the entire developmental system are innate to that system: they are determined by the resources internal to it. Both Wilson and Wimsatt entertain this view, or something close to it (Wilson, 1999, pp. 363-64; Wimsatt, 1999, p. 160), although neither embraces it. It is a challenge, though, to see where precisely it goes wrong.

It might fall to eliminativism. The advocate of HEC might argue either that innateness, in general, is too fractured a concept to be of any use (Griffiths, 2002) or that the particular conception of innateness currently at issue floats too far from the traditional one to merit the title. Such qualms miss the mark, though, by failing to appreciate the fundamental role of the internalist view in historical and contemporary thinking about innateness. According to Plato, knowledge is innate because it is already in the mind, as a result of the soul's experiences prior to earthly embodiment. On Leibniz's view, knowledge is present in the mind like veins in marble and thus is innate. Similarly for most contemporary views. The idea that what is innate is what is genetically determined expresses the internalist view: genes constitute the relevant material internal to the system, and therefore, whatever they create is innate. Various other views, e.g., that what is innate is canalized, universal, or typical to the species, arguably constitute diagnostic measures of the organism's internal contribution – the contribution of what is internal to all members of a given species and in virtue of which they are members of that species. At the very least, the internalist conception constrains theories of innateness: any conception of innateness that rules a trait to be noninnate when the system's internal properties are primarily responsible for that trait's appearance thereby faces a serious objection. It is not credible, then, to dismiss the internalist concept of innateness as unprincipled, fractured, or deviant. Thus, DST does not secure a non-nativist reading of HEC.

Moreover, closer examination of DST's implications regarding systems individuation undermines HEC's appeal to DST as a general source of support. One of the primary problems currently faced by DST is the problem of systems individuation in the theory of selection (Griffiths and Gray, 2004, pp. 423-24). Advocates of DST have made a strong case that external resources contribute significantly to the traits on which selection operates. Nevertheless, this does not settle the issue of what are, properly speaking, biological individuals; if we were to accept that anything causally relevant to the development of a trait becomes part of a biological individual, we would saddle ourselves with a profligate metaphysics for biology. There are alternatives, however. If one focuses on selection for individual traits, one can, in many cases, interpret the selection process either of two ways: as the selection for extended biological systems or as selection for traits of individual organisms that occurs within a particular environment (which might, for example, include the presence of other organisms with the same or a complementary trait; Sterelny and Griffiths, 1999, pp. 166-72).

To distinguish cases, Sterelny and Griffiths propose a common-fate criterion (1999, pp. 161, 172-77, following Wilson and Sober). Say that we encounter what appears to be an extended system, one that includes something more than a single organism. We can ask to what degree the various components of that system are subject to the same selection pressure, i.e., to what extent the reproductive fate of the components is shared. In the extreme case, every part of a system reproduces together or not at all; there is no independent reproduction or survival of parts.[8] In such cases, a trait exhibited by the entire system is possessed by that system as a biological individual. In other cases, though, the organism interacts with its environment in reproduction-enhancing ways, without shared fate. In these cases, traits are selected for because of their bearers' interactions with components of some larger system; it is only in the context of those interactions that the trait confers its selectional advantage. Still, the bearers in question can and sometimes do go it alone: they can survive and reproduce in the absence of the other components on which the utility of some of their own individual traits depends. In such cases, although an extended system might seem to exhibit a single trait, the components of that system are reproductively independent; and thus the traits on which selectional forces operate are traits of the components. In these cases, componential explanation -- i.e., explanation given in terms of selection pressures operating on the traits of the components as biological individuals -- is to be preferred; and in this way, we avoid profligate metaphysics. Many selection-based explanations depend on distinguishing parts of the system. Once such smaller individuals have been admitted into our ontology, there is no reason to add an extended system; analysis in terms of the smaller individuals situated in their environments suffices to explain whatever effects the trait in question has, including effects that change how selectional forces operate on the trait.[9]

Thus, although DST might lay firm biological ground for some extended cognitive systems, significant limitations apply: extended systems exist only where the parts of the extended system share a common fate; only in those cases do facts about extended trait selection bear on cognitive systems individuation, because only in those cases do facts about extended trait selection bear on the individuation of biological systems. Problem is, most of the systems of interest to HEC-theorists do not satisfy the shared-fate criterion; these systems exhibit significant asymmetries among their components, analogous to asymmetries resulting in reproductive independence in the biological context. In lieu of some other more convincing criterion for the individuation of biological systems, these asymmetries undermine the DST-based case for HEC.

Consider the genesis of extended cognitive systems. The volition of the organism, its intention to take up tools, and its capacities to do so are asymmetrically responsible for the creation of extended cognitive systems.[10] For example, an organism with no need to solve complex mathematical problems does not create a system of written numerals the manipulation of which facilitates the solving of those problems. I would like to pursue this concern about asymmetry, but in the interest of variety, I shall do so in the context of a different example, one presented by Clark (2003, pp. 76-77): the artist using a sketchpad who creates drawings via a feedback loop. The artist begins a sketch by making preliminary figures. The results of these early strokes impinge on the organism, causing her to "see" the artistic possibilities a new light and thus to make different, often more sophisticated sets of new strokes; the cycle repeats, with the final art-object taking a form that the artist would not have envisioned without the use of the sketchpad as a tool. Nevertheless, if the organismic subject had not been interested in drawing, the organism would not have taken up a sketch pad, i.e., the extended system in question would never have come into existence. Asymmetrically, the sketchpad's interests, goals, or other internal processes provide little impetus for the creation of the system in question. (An artist might describe her sketchpad as calling to her, but she would, I take it, be speaking figuratively of her internal states.)

Admittedly, there is some bare counterfactual sense in which the sketchbook causes the existence of the extended system: in the nearest possible world in which the sketchpad does not exist, the particular extended system in question does not come into existence. Despite this causal contribution on the part of the sketchpad, the asymmetry is genuine. Keep in mind how lopsided are the relative contributions of the organismic and extra-organismic portions of the extended system. Consider what happens if a particular sketchpad or a particular artist is deleted from history. The human, wishing to draw, is likely to find or make a different sketchpad if the particular one she would otherwise have gotten is destroyed. The sketchpad seems less likely to find or make a different human if her user has been removed. This may be partly a contingent statistical factor – at any given time the proportion of artists who want to sketch who find sketchpads is much greater than the proportion of sketchpads ready for use that in fact get used – but behind it lies the point about impetus made above: the human who wants to sketch will go out of her way to find or make a sketchpad; the sketchpad instantiates no internal processes that home in on or actively create users. Now consider a slightly different kind of counterfactual variation: type-level obliteration. Wipe the sketchpads from history and human organisms still exist. Wipe the humans from history, and there will be no sketchpads. Finally, take a special case of the preceding point, a world where no sketchpads exist or ever have existed. It is likely that human organisms will create sketchpads, but the converse will not happen: no sketchbook will bring a human into being. So, there appears to be an asymmetrical relation between the causal processes responsible for the existence of the components of the extended system.[11]

Return now to the question of whether DST, qualified by the shared-fate criterion, supports HEC. In light of the preceding discussion of asymmetries, it seems clear that the human organism is reproductively independent of the sketch pad and thus that the composite system of sketchpad-plus-human does not satisfy the shared-fate criterion. Does this concern carry over to other examples? It would appear that the standard examples of extended systems – those involving language, mathematics, external memory storage, and nautical artifacts – manifest most or all of the asymmetries discussed above. Furthermore, although the development of these tools surely affected humans' rate of reproduction, human organisms are reproductively independent of all such resources (having reproduced without them for millennia).

In the end, then, two systems-based concerns threaten the marriage of DST and HEC. First, a plausible DST validates only a narrow range of genuinely extended biological individuals; the shared-fate criterion severely limits the number of such extended systems. Second, no matter how things turn out in respect of biology, we cannot ignore the potential for mismatch between the extended individuals established by DST and those systems claimed by HEC to be extended cognitive systems. Clearly these two worries operate together: given the broad range of extended cognitive systems, the narrowness of the range of extended biological systems dims the prospects for cross-disciplinary fit. It may well be that natural forces sometimes select for extended systems, but such systems might not be the ones of interest to psychology (Rupert 2004, n22; Wilson 1999, p. 363).[12] If, for example, extended selection requires a stable environmental contribution, yet cognitive science is interested in abilities that can be exercised flexibly across a wide range of environments, danger of mismatch looms large. These concerns about mismatch, together with the nativist results of combining DST and HEC, suggest that the antinativist HEC-theorist should try a different tack: to set aside entirely the appeal to biology – DST and all the rest – and look instead to psychological criteria of innateness.

II. Extended systems, nativism, and psychology

Before turning to psychological criteria of innateness, we should address the prior question "Are there extended systems that might plausibly serve as bearers of innate psychological capacities or traits?" If there are no such systems, it is moot whether psychological criteria would pronounce their cognitive traits to be innate. To prosecute this question, it is helpful to take up more general issues pertaining to the explananda of cognitive psychology, its investigative methods, and the relation of both to HEC.

Humans categorize, perceive, remember, use language, reason, and make sense of the actions of others -- these and more are abilities of persisting systems. In contrast, most actual extended systems are short-lived: they involve the human organism's short-term use of or interaction with some kind of external resource. The importance of systems that persist and cohere, even through change, is especially clear in developmental psychology: we want to know how *that* system -- a single developing human -- came to be the way it is and how a similar course of development happens, on average, for the relatively homogeneous multitude of such persisting human systems. We want to understand how and why the capacities and abilities of individual persisting systems change over time, eventually taking a stable form. If the systems to be investigated were relatively short-lived systems -- the organism together with its immediate linguistic environment, for example -- developmental inquiry would seem incoherent. We want to be able to explain why, for example, the child categorizes on the basis of appearance at age two but pays more attention to insides at age five (Carey, 1985; Keil, 1989; Markman, 1989). How can this question be sensibly posed -- and in such a way that it might motivate a research program -- if all that exists are ephemeral systems, lacking in the continuity that makes the child at five the same system as the child at two?[13]

In response, the advocate of HEC might remind us of her aim to reconceptualize cognitive systems. Such reconceptualization founders on a dilemma, however: either we pursue genuine reconceptualization of cognitive systems as relatively short-lived systems, at great cost to cognitive psychology, or we can jury-rig a method of cognitive systems individuation that preserves the successes of cognitive psychology and is consistent with a viable investigative method -- but is unmotivated and unnecessarily complex.

Consider what is lost if the HEC-theorist pursues the former tack. The typical experiment in cognitive psychology yields useful and coherent results by assuming such privileged grouping of various short-lived systems (i.e., the typical experiment presupposes that subjects are persisting, organismically bound cognitive systems).[14] Think of the multitude of within-subject analyses of results on short series of experiments, all data lost, absent privileged groupings. Radical reconceptualization, however, offers developmental psychologists no more reason to be interested in, for example, the series of temporal segments we normally associate with Sally from ages two-to-six than it offers to be interested in, say, Sally, aged two, together with a ball she was bouncing on some particular day, Johnny, aged five, together with the book he was reading on some particular afternoon, and Terry, aged seven, plus the stimulus item he has just been shown by an experimenter. It is simply not clear how one should proceed after giving up the traditional method of systems individuation.

These problems are not limited to developmental psychology. Investigations of adult capacities, for memory and language use, for example, normally presuppose that the subjects of investigation are individual persisting systems. Some such studies are explicitly longitudinal (Bahrick, 1979, 1984), and thus much like developmental psychology in the relevant respects. Beyond these cases, psychologists and linguists frequently study contextual effects. It is striking that the same person behaves in one way in one context -- say, when not primed -- and behaves differently in another slightly different context -- when, in contrast, she has been primed. There is an enormous body of fascinating literature filled with experiments interpreted in just this way, their explananda taken to be persisting individuals having various capacities or abilities that they exercise in different ways in different contexts.

A focus on the persisting individual is also evident in research on perception. We would like to know why, for example, the subject perceives certain features under some conditions – say against a particular background – but does not perceive those same features under other conditions (see, for example, various results surveyed in Treisman, 1998). The experimenter asks a single system to perform various visual tasks, and the outcome sheds light on the process by which that system sees. Perhaps, as is sometimes emphasized, the subject *does* something as a way of getting information visually (Churchland, Ramachandran, and Sejnowski, 1993). Still, such results reveal something about the ability of a single persisting system: they reveal how *that* system gets visual information.[15]

The preceding argument against extended individuation might seem to have ignored an important aspect of standard methodology: researchers frequently assign experimental subjects to different groups; in a typical experiment, these consist of a control group and an experimental group. In such experiments, researchers do not appear to be investigating the capacities of individually persisting systems as they change over time or as they exercise their capacities in varying circumstances. This rejoinder misinterprets standard methodology, however. Researchers assign subjects to different groups on the assumption that the set of members of each group represents a standard distribution of cognitive skills and capacities across a population of members of the same kind. By statistical analysis of the results, we think we discover something about the way the *standard persisting human system* reacts under different conditions.

The HEC-theorist might, in response, point out that data is sometimes analyzed simply by condition or by question. This rejoinder, however, takes too narrow a view of the ways in which analysis by question is used. Normally, researchers compare the results of analysis by condition to the results of a similar analysis of results in conditions – the point being to see how systems of the same kind behave under different conditions. Such comparative analysis presupposes that the individuals involved in both (or the many) conditions are representative of the human population; thus, we are, once again, comparing the responses of a single (kind of) system under different conditions.

Of course, sometimes we naturally and legitimately group together temporally disjoint systems. Consider the practice of medical doctors, who talk about the same patient over time, even though, at the biological level, that patient changes his constitution (cells die, new ones form). This, however, only emphasizes the need for a principle of organization to ground the groupings. Organismic integrity and the way it physically grounds health and disease recommend treating a person over time as a single patient, returning for visits. Insofar as an organizing principle motivates the groupings of alleged cognitive systems into privileged sets (and legitimates the chosen groupings), this principle is similarly organism-based -- which is, of course, just how traditional cognitive science identifies its systems of interest. This moderate position – the second horn of the dilemma mentioned above – is not worth its price. The HEC-theorist buys a highly counterintuitive claim about minds (that they are extended beyond the organismic boundary) for the cost of unnecessary complications in, without any substantive departure from standard individuative practice.

We should not pronounce in advance what a completed cognitive psychology will bring. Nevertheless, insofar as we can make out a genuinely extended-systems-based alternative to existing methodology, it faces deep problems. It introduces a profligate set of distinct cognitive systems the richness of which confounds standard methodology, with no productive replacement in sight. In contrast, it is open to the HEC-theorist to partition extended systems into useful subgroups, but this amounts to little more than a co-opting of the success of standard methodology. On measures of simplicity and conservatism, then, this strategy clearly loses out to the traditional approach (and to an embedded approach). Of course, costly revisions in theoretical frameworks are justified when they offer sufficient gains along other dimensions, e.g., in explanatory power or accuracy. The shift under consideration does not, however, do so. The range of provocative and fruitful results in contemporary cognitive science can reasonably and manageably be cast in terms of organismically bounded cognitive systems that frequently interact intimately with their environments.

Where does this leave us with respect to HEC and nativism? Above I recommended that HEC abandon biology-based theorizing about innateness and draw instead from sciences closer in subject matter to HEC's own domain. Before concluding this section, let us briefly consider a pair of widely discussed psychological criteria of innateness. First, take the primitivist view (Cowie, 1999; Fodor, 1981; Samuels, 2002). According to this view, a psychological trait is primitive if and only if there is no psychological account of its appearance.[16] The asymmetries discussed above suggest that in most cases, the appearance of an extended cognitive trait has a psychological explanation: extended systems, and thus their traits, originate in the psychology of the organism and thus would not count as innate by the primitivist criterion.[17] Application of a second commonly used psychological measure of innateness, domain-specificity (Cowie, 1999;

Keil, 1990), yields similar results. The basic idea is that if a cognitive capacity can only be exercised with respect to a proprietary set of inputs or tasks, it is innately dedicated to those inputs or to that task-domain, and the capacity in question is thus innate. Many of the relevant capacities of extended systems are not domain-specific. Pen and paper are used to draw diagrams, solve math problems, and make sketches, and thus, will not be categorized as innate by this criterion. Surprisingly, this might even hold in the case of extended visual systems. Keep in mind that the organism together with what it visually perceives constitutes the relevant system; so, even if the organism's visual system is domain-dedicated, it is not clear what are the domain-specific restrictions relative to the *entire* extended system – i.e., the visual apparatus taken together with what is seen.

Thus, the antinativist HEC-theorist wins a Pyrrhic victory. The psychological criteria of primitiveness and domain-specificity plausibly categorize at least some of the relevant cognitive capacities in a way consistent with the non-nativist leanings of many situated theorists. Consequently, *if* extended systems were the right sort of thing to which to attribute psychological properties, the advocate of HEC would have found antinativist support in the application of these psychological criteria for innateness. Cognitive science, however takes as its explananda cognitive capacities of persisting systems; it is wrong-headed to attribute innate capacities to systems that are short-lived relative to the capacities in question, largely because the capacities of the short-lived systems are not the proper objects of inquiry in cognitive science. Thus, I conclude that there are few or no innate capacities of extended systems – but for reasons HEC-theorists cannot embrace.

III. Embedded cognition and context-specific representations

Many embedded views appeal to context-specific representations, the contents of which seem best understood in non-nativist terms. These context-specific representations amount to little more than mental demonstratives, internal placeholders that take ephemeral values relative to the task being performed (Ballard, Hayhoe, Pook, and Rao, 1997); as such, their *contents* depend heavily on the contribution of the environment. The embedded view's deep reliance on context-specific representation promises to provide the situated program with an importantly non-nativist aspect, for representations play a central role in many embedded models, and questions about representation, concepts, and content stand at the center of many debates about nativism (Cowie, 1999; Fodor, 1981; Rupert, 2001). If embedded cognition avoids nativism along this dimension of the debate, embedded cognition is non-nativist in at least one substantive respect. Nevertheless, although there is a fairly robust sense in which the content of context-specific representations is not innately determined, the theory of content best suited to such representations entails a significant nativist contribution to their content – or so I argue in this section.

Generally speaking, the embedded approach aims to minimize the amount of internal representation used to model the human performance of cognitive tasks (Ballard et al., 1997; Clark, 1997, chapter 8; McClamrock, 1995, chapter 6).[18] To appreciate the kind of minimization at issue, let us first consider a contrasting approach. On a more traditional understanding of cognition, the subject performs a complex analysis of a problem before acting: in the case of action or inference based on visual perception, such complex analysis involves the construction of a detailed, internal representation of the immediate environment; in the case of abstract planning (e.g., the choice of which college to attend), the subject might explicitly represent the costs and benefits of each of a range of options, as well as the likelihood of success of each option and the maximum cost the subject is willing to pay to achieve her goal. On an embedded view of the former case, the subject uses context-specific correlations to act on the basis of visual perception in a way that requires substantially less in the way of representational resources than would be required to build a detailed internal representation of the relevant environment. To illustrate, consider Clark's (1995) example of finding the photo-development counter at a super-market. Given the market-dominance of Kodak, the area above or around the photo-development counter is normally splashed with yellow. This suggests a simple strategy to the consumer in search of the photo counter: enter the super-market, and swivel one's head about looking for a large patch of yellow.

At work in this example are three tactical principles by which embedded models minimize the subject's use of internal representational resources. First, the subject need not explicitly represent – in working memory or any other sort of cognitive workspace – any very elaborate theory or conception of photo counters. She needs only to activate a routine that exploits the local, contingent correlation between patches of yellow and the location of photo counters. Second, the subject represents what might be called a 'coarse-grained' property: the subject looks for a large patch of yellow, not one of any particular size. Third, the subject collects only the information she needs. Mind you, the requisite representational resources far exceed nil. The subject must, for example, represent in some way the project or plan she is engaged in: a person should not wander through life looking for yellow patches. Once the goal has been determined, though – e.g., to get one's photos developed – and the subject has found and entered the store, the amount of information needed, at that moment, is fairly small. The subject need not know the general layout of the building, the location of the restrooms, the number of cashiers on duty, or which cashier lines are open. The subject might need access to some of these other facts at some time during her visit; but when she enters the store and begins looking for the photo counter, she can ignore these other matters.

A more general idea binds these three principles to the embedded theorist's minimalist approach to representation: as much as possible, let the environment do the cognitive work (Brooks, 1991; Clark, 1989, p. 64). Perhaps surprisingly, all of these principles – from the overarching idea of letting the world do the work to the more specific tactical principles particularly to do with representation – apply also to the case decision-making, where many of the points about visual engagement with the world might seem to have little application. The general strategy of letting the world do the work could be applied in the following way: the agent sets a maximum acceptable cost of achieving her goal as well as a minimum probability of success, then accepts the first satisfactory option offered to her by the environment. This requires *some* internal calculation on the subject's part and thus some internal representation; but it is much less demanding than an optimizing comparison of a good number of options. Thus, this approach remains in step with the embedded emphasis on minimizing internal representational resources. Furthermore, the three principles listed in connection with the example of visual cognition are at work, at least to some extent. In many actual environments, there is a contingent correlation between *being an option that presents itself* and *being a satisfactory option*. Of course, if there were only a weak correlation or none at all, the subject might spend a lot of time performing a series of quick and dirty analyses of the many possibilities that present themselves, determining serially that none will do. It is often observed, though, that humans structure their environments, institutions, and social interactions in such a way as to strengthen the kind of correlation at issue – so that, for example, if the subject decides on the first college she comes across that meets her minimum standards, she will probably achieve her related goals (for example, to become reasonably well educated). Also, given that the subject makes no detailed comparison of options, she can represent coarse-grained properties – such as *being a fairly prestigious school* – in place of any attempt to make a fine-grained determination of, for example, relative degree of prestige of a range of alternatives. Imagine a case where two or more options fall in roughly the same ballpark in terms of costs and benefits; if the subject's charge is to figure out which is better overall, she cannot coarse-grain – coarse-graining leads to a tie. Decision-making under such conditions required fine-grained comparison. If, however, the subject considers the first opportunity that presents itself, sees that it offers sufficient promise of success while falling at or below the maximum acceptable cost, then she is off to the races. The discussion of coarse-graining makes clear that embedded accounts of decision-making embody the last of the three representation-minimizing principles discussed in connection with visual perception: collect only the information needed. The amount of information one needs to gather to determine that a single school is fairly prestigious falls far short of the amount the subject would need to gather to determine which of a number of fairly prestigious schools beats all the others in the group.

The drive to minimize the use of internal representational resources seems to recommend heavy reliance on *ad hoc*, context-specific representations. Such representations do not persist, content intact, but rather depend for their content on the immediate and typically short-lived contribution of their environment. The embedded approach would thus seem to suggest a cognitive system that does as much as possible with only a standing collection of mental demonstratives (pointers instantiated in visual cortex, for example) that take different values depending on their short-lived relationships to environmental factors; or perhaps in the extreme, the subject might construct new pointers in new situations, so that even individuated *nonsemantically*, the pointers have little integrity over time. On the embedded view, then, representational content clearly depends on the contribution of the subject's immediate local environment, something that, having set aside HEC and DST, we can assume is not innate to the subject. Let us pursue this approach to representation, to see whether one can reasonably expect it to deliver a robustly non-nativist understanding of mental content.

It might be too demanding to insist that the embedded theorist have in pocket a fully worked out theory of content for context-specific representations. Furthermore, limitations of space prevent a sustained attempt to determine here which, among the various leading theories of mental content, correctly characterizes the kind of content carried by context-specific representations.[19] Of extant theories, Robert Cummins's (1996) combination of intenders and isomorphism can be most naturally applied to the embedded approach. On Cummins's view, the cognitive system consists largely of functionally discrete components – intenders – each of which aims at a particular kind of target, for example, the spatial layout of the immediate environment. Each intender employs a set of representational primitives that can be activated in complex arrangements. The representational value of each arrangement is determined by isomorphism alone, leaving each arrangement wildly ambiguous. Nevertheless, a mental state, a belief, for example, can be determinately correct because the function of an intender in which the representation appears limits that representation's contribution to something from the appropriate range of targets (i.e., targets appropriate to the intender in which the representing structure is activated).[20]

How does any of this bear on the question of innateness? It depends partly on how we resolve a question that Cummins himself leaves open, the question of how the intenders acquire their functions. On the one hand, Cummins considers the approach he helped to found, that of homuncular or systemic decomposition. According to this view, the contribution of a component to the performance of its containing system determines that component's function. If intender *I1* contributes what – in the context of the system's overall behavior – is best construed as perceptual uptake, then *I1* has some particular aspect of the immediate environment as its target (which aspect depends on what kind of perceptual uptake *I1* is best understood as contributing to the overall performance of the system). Cummins also considers a Millikan-style teleofunctional view (Millikan, 1984). On this approach, evolutionary selection determines an intender's function (Cummins, 1996, chapter 8).

The former approach is problematic, for reasons parallel to those Cummins (1996, pp. 41-51) wields against functionalist (or conceptual role) theories of content. In conflict with a fundamental constraint on any theory of mental content, functionalist theories do not seem to allow for error. Take a case in which we think a subject has made a representation-based error. Recall that the functions at issue are computational-mathematical functions; thus, it will always be possible to locate some other function – i.e., some function other than the one we thought the system was computing – relative to which the system's computation was accurate. Similarly, though, this strategy of alternative interpretation applies to any case of what we might think of as erroneous representation within an intender. If, for example, it seems that *I1* contributes information about spatial layout to the functioning of its containing system, but on some particular occasion its output causes the containing system to trip over a table, we can reinterpret *I1*'s function as representing the spatial layout of the room adjusted in such-and-such a way (e.g., *I1* is supposed to represent spatial layout excepting tables on Tuesdays, if that is when the trip occurred).

The teleofunctionalist view fares better because, presumably, it is a matter of fact what a particular functional component was selected for (assuming it was selected for the performance of a function!). If a component kept its owners alive longer because it conferred the ability to see berries, then this component had better have the function of accurate visual detection. There is no case to be made that accurate berry detection on Tuesdays led to reproductive advantage, partly because there is no relevant aspect of the organism's physiology connected nomically to the property *being eaten on a Tuesday*. But – and here is the punch line – if embedded cognition relies on the teleosemantic version of Cummins's theory of content, content is innate in one very important respect. Any given representing structure is isomorphic to a great number of external structures; the intender plays an essential disambiguating role by determining which of a representation's many values that representation contributes to the content of the mental state in question. The representational content of that mental state rests immediately and synchronically on the function of the intender in which the representation appears, and the function of that intender is determined innately.

One might think that causal contact with the environment minimizes nativist commitments: after all, in the typical embedded model, the subject causally interacts with the portion of the world to which her context-specific representations are bound. Even here, however, innate contributions plays a prominent role. In one of Ballard et al.'s (1997) examples, a single pointer gets bound to the color of a block in a visual display, but that block has texture as well as shape and other visually detectable properties. What determines that the pointer is bound to the *color* of the block? Presumably, the pointer's content depends on the function of that pointer in a routine (which is partly determined by the intender(s) involved, on Cummins's view). This routine can have complex structure, for example, in terms of the way it uses the bound information, and this structure determines to what the pointer is bound.[21] To the extent that this complex structure of the routine arises from its teleofunction, its contribution to the content of the relevant cognitive state is innately determined to a significant extent.

Of course, there is also a significant sense in which embedded models remain empiricist: the world contributes specific content, e.g., *green* or *yellow*, to the pointer. Here Wilson's (2004) two-dimensional analysis of the nativism-empiricism debate may be useful. Put simply, there are two dimensions of complexity, the internal and the external. If the internal mechanisms driving some cognitive process are rich and complex and largely independent of environmental control, while the external contribution is minimal, we rightly label that process 'innate'. If, in contrast, the internal mechanism driving a cognitive process is simple in structure, while the external contribution is rich and complex, an empiricist conclusion is warranted. In mixed cases, such as the present one, the two-dimensional analysis fails to deliver a particular judgment – innate or not. Thus, Wilson's two-dimensional analysis explains our discomfort in issuing either a firmly nativist or firmly empiricist judgment about the content of context-specific representations.

IV. Nativism and persisting representations

Theories of cognition must make some allowance for persisting, internal representations. Children employ amodal representations from early on; concepts are used in abstract thought, when one is, for example, alone in the study; theorists conceptualize, reason, design experiments, and write papers (about, for example, embedded cognition); and what is more, embedded models themselves appeal to standing routines, intentions, or programs in ways touched on above. I do not offer this observation in argument against embedded models. Rather, the present section fleshes out a sense in which embedded models help to explain – in an empiricist way, no less – how the content of persisting representations is fixed.

Think of the human cognitive system as two-layered; this is surely an oversimplification, but it allows us to focus on the relevant issues without, I think, distorting the conclusions reached. One layer consists of perceptual systems or modules. Call the representations employed at this level ‘peripheral representations’. These include such representations as the visual image of a tree currently in the subject’s view; typically the context-specific representations posited by embedded models fall into this category. The second layer is what some theorists call ‘central processing’ (Fodor, 1983). Let us dub the representations appearing at this level ‘central representations’. Theoretical, abstract, and nonsensory concepts, such concepts as PROTON, JUSTICE, and CHAIR (as a general kind) fall into this category.

It might appear that central representations come into no special contact with the embedded program. I propose a contrasting view: peripheral representations mediate the causal relations that fix the content of many central representations; furthermore, one of the most commonly cited examples of embedded representational activity – the active use of language – contributes significantly to the fixation of the content of central representations, by helping to bring them into the causal relations that determine what they represent. This view does *not* require that central representations *inherit* their content directly from some selection of peripheral representations or from external symbols; neither does it require that peripheral representations or external symbols *define* central representations. Thus, the current proposal is not empiricist in the traditional sense: central representations can and typically do have content that outruns the content of the subject's sensory representations or logical constructions from those sensory representations.[22]

In what sense, then, is the current proposal empiricist? Because atomic central representations are stable, reappearing units, it seems likely that their content can be fixed without the contribution of an intender that determines the target of the representation on particular occasions of use. Moreover, application of the primitivist criterion suggests that the content of central representations is not innate; peripheral representations and external symbols mediate content-fixation, which typically involves a fairly detailed psychological explanation of the fixation of the content of central representations (cf. Cowie, 1999). This non-nativist implication of the primitivist criterion follows from a central commitment of the embedded approach. A recurring theme in the embedded literature has been the extent to which humans simplify their cognitive tasks by structuring their external environment in ways that facilitate the performance of those tasks (Clark, 1997; Dennett, 1996; McClamrock, 1995). Some such simplification can be effected through the use of language and the assignment of a privileged role to certain observations. Straightforward examples involve the construction of written descriptions or a visual diagrams meant to convey a new concept to others, i.e., to get their central representations causally connected to the correct properties or kinds. In another sort of case, the individual formulates her own observation-based heuristic, for example, of the form "when I see such-and-such, it means there is a *C* present," where *C* is a central representation. In all of these cases, a peripheral representation comes to mediate the activation of a newly created central representation or a central representation to which the peripheral representation had not previously been connected. In the initial case, where the subject coins a new central representation that representation might not carry any determinate content or at least not the content it will ultimately carry (Margolis, 1998; Rupert, 1996, chapter 4, 1998/1999, 2001). As various peripheral representations become causally connected to the newly coined central representation, the latter can come into a content-determining causal relation to an external property, kind, or individual. Adverting to the human habit of actively structuring her environment so that she has certain sensory experiences constitutes a powerful psychological explanation of how a human comes into the right causal position to have a central representation with the content *electron*. (In fact, in the case of some kinds, for example, *electron*, this process seems the only plausible way for a human to get a representation with that kind as its content – see Fodor, 1998.)

Granted, the contents of peripheral or organismically external representations do not exhaust the contents of the central representations in question, and thus, the organism must make some internal contribution to the fixation of content for the latter representations. For Fodor, this contribution – together with the absence of a hypothesis-testing explanation of the acquisition of central representations – secures the innateness of those representations (Fodor, 1975, 1981). To a certain extent, I think Fodor is right. Even with a set of mediating mechanisms in place, including peripheral representations, the organism must at least be disposed to come into the appropriate content-fixing causal relation to one external kind or property rather than another. Thus, there is *some* innate contribution to the content of central representations. All the same, there is a rich and complex psychological explanation of how mediating mechanisms contribute to the fixation of content for central representations, while the nonpsychological story is very thin. Furthermore, both the nature of the environment and the causal properties of the material available to instantiate peripheral representations in the human brain substantially limit the innate, internal contribution. There is a limited number of genuine properties and kinds present in the subject's environment. Moreover, the mechanisms that mediate the tokening of the central representations in question themselves have limited causal properties: the visual cortex may only be capable of relaying reliable signals to central processing about certain of the available properties among that limited set present in the environment. Thus, it may be that the human mind develops a representation of, e.g., *electron* in a given set of circumstances not because a central representation has *electron* as its innate content but rather because the contingent facts about the environment the human is thrown into and the physical substance its perceptual mechanisms are made of do not allow the central representation in question to be attached in the proper fashion to any kind other than *electron*. It is plausible that the physical mechanisms realizing sensations were selected because of the way they respond causally to certain of the physical features of the subject's environment. In contrast, the physical material that realizes central processing was selected for its relative flexibility, not its physical ability to carry information about any specific aspects of the environment. Thus, central processing contributes very little to the content of such concepts as ELECTRON that is innately keyed specifically to their content.

In the end, we might try grafting Wilson's two-dimensional view to the primitivist theory of innateness – the two dimensions becoming psychological and nonpsychological contributions to the development of the trait (acquisition of the concept) in question. In the present case, this approach yields a fairly strong judgment on the empiricist end of the spectrum. The psychological story of content fixation is quite rich, whereas the nonpsychological internal contribution consists only in a disposition to be causally connected in one way rather than another given a certain complex psychological and physical environment.[23]

Much of this chapter has been an exploration of possibilities, an attempt to show that the situated program does not reside squarely in either the empiricist or nativist camp. Along the way, I have argued that, on various plausible assumptions, some situated models are subject to clear nativist gloss, contrary to the leanings of many situated theorists. As development of situated program proceeds, we should keep in mind the wide range of theoretical possibilities open in respect of nativism and the situated modeling of cognition.[24]

References

Ariew, A. (1999). Innateness is Canalization. In Hardcastle, 1999 (pp. 117-38).

Bahrick, H. P. (1979). Maintenance of Knowledge: Questions about Memory We Forgot to Ask. *Journal of Experimental Psychology: General, 108*, 296-308.

Bahrick, H. P. (1984). Semantic Memory Content in Permastore: Fifty Years of Memory for Spanish Learned in School. *Journal of Experimental Psychology: General*, *113*, 1-29.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., and Rao, R. P. N. (1997). Deictic Codes for the Embodiment of Cognition. *Behavioral and Brain Sciences*, 20, 723-42.

Block, N. (1981). Introduction: What Is Innateness? In N. Block (Ed.), *Readings in the Philosophy of Psychology*, vol. 2 (pp. 279-81). Cambridge: Harvard University Press.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge: MIT Press.

Brooks, R. (1991). Intelligence without Representation. *Artificial Intelligence*, *47*, 139-59.

Butler, K. L. (1998). *Internal Affairs: Making Room for Psychosemantic Internalism*. Boston: Kluwer.

Carey, S. (1995). *Conceptual Change in Childhood.* Cambridge: MIT Press.

Carruthers, P. and Boucher, J. (Eds.). (1998). *Language and Thought: Interdisciplinary Themes*. Cambridge: Cambridge University Press.

Churchland, P. S., Ramachandran, V., and Sejnowski, T. (1994). A Critique of Pure Vision. In C. Koch and J. Davis (Eds.), *Large-Scale Neuronal Theories of the Brain* (pp. 23-60). Cambridge: MIT Press.

Clark, A. (1989). *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing*. Cambridge: MIT Press.

Clark, A. (1995). Moving Minds: Situating Content in the Service of Real-Time Success. In J. Tomberlin (Ed.), *Philosophical Perspectives*, vol. 9: *AI, Connectionism, and Philosophical Psychology* (pp. 89-104). Atascadero, CA: Ridgeview.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge: MIT Press.

Clark, A. (1998). Magic Words: How Language Augments Human Computation. In Carruthers and Boucher, 1998 (pp. 162-83).

Clark, A. (2003). *Natural-Born Cyborgs*. Oxford: Oxford University Press.

Clark, A. (2005). Intrinsic Content, Active Memory, and the Extended Mind. *Analysis*, *65*, 1-11.

Clark, A., and Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*, 7-19.

Clark, A., and Thornton, C. (1997). Trading Spaces: Computation, Representation, and the Limits of Uninformed Learning. *Behavioral and Brain Sciences*, 20, 57-66.

Cowie, F. (1999). *What's Within? Nativism Reconsidered*. Oxford: Oxford University Press.

Cummins, R. (1996). *Representations, Targets, and Attitudes*. Cambridge: MIT Press.

Dennett, D. (1987). *The Intentional Stance.* Cambridge: MIT Press.

Dennett, D. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.

Dretske, F. (1998). *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT Press.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge: MIT Press.

Fodor, J. (1975). *The Language of Thought*. Cambridge: Harvard University Press.

Fodor, J. (1981). The Present Status of the Innateness Controversy. In J. Fodor, *Representations* (pp. 257-316). Cambridge: MIT Press.

Fodor, J. (1983). *The Modularity of Mind*. Cambridge: MIT Press.

Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge: MIT Press.

Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge: MIT Press.

Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Fodor, J. (2003). *Hume Variations*. Oxford: Oxford University Press.

Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge: MIT Press.

Goldin-Meadow, S. and Zheng, M. (1998). Thought before Language: The Expression of Motion Events Prior to the Impact of a Conventional Language Model. In Carruthers and Boucher, 1998 (pp. 26-54).

Griffiths, P. (2002). What is Innateness? *The Monist*, *85*, 70-85.

Griffiths, P. and Gray, R. (2004) The Developmental Systems Perspective: Organism-Environment Systems as Units of Development and Evolution. In K. Preston and M.

Pigliucci (Eds.), *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes* (pp. 409-31). Oxford: Oxford University Press.

Griffiths, P. and Stotz, K. (2000). How the Mind Grows: A Developmental Perspective on the Biology of Cognition. *Synthese*, *122*, 29-51.

Grush, R. (1997). The Architecture of Representation. *Philosophical Psychology*, *10*, 5-23.

Hardcastle, V. G. (Ed.) (1999). *Where Biology Meets Psychology: Philosophical Essays*. Cambridge: MIT.

Haugeland, J. (1995). Mind Embodied and Embedded. In Y. Houng and J. Ho (Eds.), *Mind and Cognition* (pp. 3-37). Taipei, Taiwan: Institute of European and American Studies, Academia Sinica.

Houghton, D. (1997). Mental Content and External Representations. *Philosophical Quarterly*, *47*, 159-77.

Hurley, S. (1998). Vehicles, Contents, Conceptual Structure, and Externalism. *Analysis*, *58*, 1-6.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge: MIT Press.

Keil, F. (1989). *Concepts, Kinds, and Cognitive Development*. Cambridge: MIT Press.

Keil, F. (1990). Constraints on Constraints: Surveying the Epigenetic Landscape. *Cognitive Science*, *14*, 135-68.

Margolis, E. (1998). How to Acquire a Concept. *Mind & Language*, *13*, 347-69.

Markman, E. (1989). *Categorization and Naming in Children: Problems of Induction*. Cambridge: MIT Press

McGinn, C. (1989). *Mental Content*. Oxford: Basil Blackwell.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge: MIT Press.

McClamrock, R. (1995). *Existential Cognition: Computational Minds in the World*. Chicago: University of Chicago Press.

O'Regan, J. K. (1992). Solving the 'Real' Mysteries of Visual Perception: The World as an Outside Memory. *Canadian Journal of Psychology*, *46*, 461-488.

Prinz, J. (2000). The Duality of Content. *Philosophical Studies*, *100*, 1-34.

Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes*. Cambridge: Cambridge University Press.

Rupert, R. (1996). The Best Test Theory of Extension. Ph.D. dissertation, University of Illinois at Chicago.

Rupert, R. (1998/1999). On the Relationship between Naturalistic Semantics and Individuation Criteria for Terms in a Language of Thought. *Synthese*, *117*, 95-131.

Rupert, R. (1999). The Best Test Theory of Extension: First Principle(s). *Mind & Language*, *14*, 321-55.

Rupert, R. (2001). Coining Terms in the Language of Thought: Innateness, Emergence, and the Lot of Cummins's Argument against the Causal Theory of Mental Content. *Journal of Philosophy*, *98*, 499-530.

Rupert, R. (2004). Challenges to the Hypothesis of Extended Cognition. *Journal of Philosophy*, *101*, 389-428.

Samuels, R. (2002). Nativism in Cognitive Science. *Mind & Language*, *17*, 233-65.

Sterelny, K. (2004). Externalism, Epistemic Artefacts and the Extended Mind. In R. Schantz (Ed.) *The Externalist Challenge: New Studies on Cognition and Intentionality* (pp. 239-54). Berlin: de Gruyter.

Sterelny, K. and Griffiths, P. (1999). *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press.

Thelen, E. and Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge: MIT Press.

Treisman, A. (1998). The Perception of Features and Objects. In R. D. Wright (Ed.), *Visual Attention* (pp. 26-54). Oxford: Oxford University Press.

Varela, F., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press.

Wilson, R. (1999). The Individual in Biology and Psychology. In Hardcastle, 1999.

Wilson, R. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences*. Cambridge: Cambridge University Press.

Wimsatt, W. (1999) Generativity, Entrenchment, Evolution, and Innateness: Philosophy, Evolutionary Biology, and Conceptual Foundations of Science. In Hardcastle, 1999 (pp.139-79.

Notes

[1] See Elman, Bates, Johnson, Karmiloff-Smith, Parisi, and Plunkett (1996, pp. 107-8) for a substantial list of references to influential nativist work.

[2] Although Clark (1997, chapter 2) discusses Thelen and Smith's developmental work in some detail, he avoids the question of nativism (but see Clark and Thornton, 1997, p. 63, where antinativist leanings are made a bit more explicit). Similarly, Varela, Thompson, and Rosch (1991, chapter 9) address biological issues of genetic determination – advocating positions that are often taken to support antinativist views – without addressing the question of nativism head on.

Note that I use 'antinativist', 'non-nativist', and their cognates as umbrella terms to cover all manner of empiricist and eliminativist views, where an eliminativist view holds that the empiricist-nativist distinction is incoherent or for some other reason should be excluded from theoretical discourse (Griffiths, 2002).

[3] For a contrasting view, see McGinn (1989, pp. 24-26, 46, 116, 210).

[4] I do not wish to oversell this point, however. Kim Sterelny has rightly emphasized the prevalence of cases in which traits alleged to be exhibited by extended cognitive systems – mathematical abilities involving the use of external media, for example – emerge only with great difficulty and often do not appear at all in the extended systems in question, even after extensive tutelage (Sterelny, 2004). Note, though, that the HEC-theorist incurs a cost by pursuing this response to the nativist "charge." She must be prepared to assert that in cases where, for example, a student cannot seem to acquire the relevant mathematical skill, an extended cognitive system exists nonetheless, so that the system's failure to develop the ability in question counts

against the canalization of that ability in the kind of system in question. This creates some tension when conjoined with the examples typically used to motivate HEC, for such examples almost invariably involve *successful* activity: showing that an extended system successfully performs the tasks we associate with human cognitive capacities is meant to show that the extended system *has* cognitive capacities, i.e., that it is an extended cognitive system. The organism's bare disposition to *try* to use pencil and paper to solve math problems does not sufficiently motivate HEC, although the antinativist HEC-theorist must count this system as an extended cognitive system if she is to avail herself of the antinativist defense bruited above.

[5] Keep in mind, however, that Wimsatt sometimes leans toward eliminativism, describing generative entrenchment as a successor concept to innateness (1999, p. 139; also see pp. 162-63 and note 22).

[6] Developmental systems theory has played an important role in discussions of group selection (Sterelny & Griffiths, 1999), where conspecifics or their traits constitute the relevant environment. Throughout the discussion of DST and HEC, I focus instead on systems comprised of individual organisms together with nonorganismic aspects of the environment.

[7] Griffiths and Stotz (2000) present their account of extended inheritance in this way; see pp. 34-39 for a juxtaposition of the extended view of selection and their antinativist reading of it. Also see Griffiths and Gray (2004, p. 425) for the view that extended inheritance is inimical to nativism. Griffiths and Stotz (Griffiths, 2002, Griffiths & Stotz, 2000) object to the association of innateness with human nature (or species nature), and this motivates some of their resistance to the nativist characterization of extended inheritance. In the present case, however, it seems that extended systems do have a nature: if there are mechanisms – both internal and external to

the organism – in place for reliably reproducing extended systems, then the requirements for the entire system's being a natural kind seem to have been met. I am, however, getting ahead of myself.

[8] The form of independence I have in mind is not full-blown probabilistic independence but rather is the more intuitive idea of there being a non-negligible probability of one thing's reproducing successfully when the other does not.

[9] This point might be articulated in terms of trait groups. A trait group in respect of trait *F* is a collection of individuals such that instantiation of *F* by some sufficient number of the individuals enhances the fitness of each of the individuals in the group. (The idea is that sometimes the presence of a trait enhances the fitness of an individual, even where that individual does not possess the trait in question but rather benefits from other individuals' possession of it.) A given individual, however, can be a member of many trait-groups the members of which vary significantly. Thus, to avoid a Byzantine network of partially overlapping, composite biological individuals (and to ensure that our account of composite, biological individual tracks only the theoretically important individuals – see Griffiths and Gray, 2004, p. 423), we might lay down the following necessary condition: for two individuals to be members of the same composite biological organism, both should be members of all, or nearly all, of the same trait groups. A given person's brain and heart, for instance, are part of the same composite biological individual because the brain's and the heart's probabilities of contributing to the appearance of others of their own kind are enhanced by the same (or very nearly same) group of traits, establishing what is, extensionally, the same trait group relative to the many traits that affect the brain's and heart's fitness. Any composite biological individual satisfying this criterion is such that its component

individuals have a very high probability of a shared fate; they will not reproduce independently of each other.

[10] Cf. Wilson's 'locus of control' argument for the privilege of the organismic system (2004, pp. 197-98; see Butler, 1998, for a similar argument).

[11] Compare Clark's sustained argument (2003) that humans are, by their nature, tool users. Clark may well be correct about human nature, and there is at least a clear sense to his claim. It is a bit difficult to understand the converse claim, i.e., that it is the nature of external resources that they be used by humans; insofar as it does have clear meaning, it seems obviously false: it is not the nature of iron ore that it be wrought by humans into automobile parts or anything else. (Note that Clark sometimes wields his argument about human nature and tool-use to a different end: to support HEC. So far as I can tell, the argument faces serious difficulties, partly because it highlights significant asymmetries between the contributions of various portions of allegedly extended cognitive systems. Given the asymmetric contribution of human organisms to the formation of such systems, I take Clark's argument to support the embedded approach rather than HEC.) See also Sterelny (2004), where he argues for other substantial asymmetries, both epistemic and representational.

[12] With regard to biology alone, different interests might lead to different principles of systems individuation (Griffiths and Gray, 2004, pp. 419-20), as well as to different standards of innateness: Ariew (1999, p. 120) argues that within biology itself, different research programs have made use of conflicting conceptions of innateness.

[13] Clark and Chalmers's (1998, p. 10) discussion of portability as a criterion for cognitive states seems partly motivated by a concern for persisting systemic integrity. There is a gap, however,

between the various hypothetical systems discussed by Clark and Chalmers – systems that may well qualify as cognitive systems with persisting abilities – and systems subject to cognitive scientific inquiry; I am concerned with the latter systems.

[14] Even those researchers who focus on short-lived, extended dynamical systems must take the traditional approach in certain respects. For example, Thelen and Smith (1994, pp. 288-89) attempt to explain children's performance on specific A-not-B trials by adverting to characteristics of a short-lived dynamical system; but insofar as Thelen and Smith's dynamical approach can explain the child's developmental trajectory – how her competency changes over time – they must appeal to changes in the capacities of the persisting organismic system; after all, the toys and the wells need not change at all. What changes is the cognitive capacity of the organism. In Thelen and Smith's terms, what changes is the child's ability to enter into dynamic relations with the environment; in which case, we should, for simplicity's sake, appeal to this organismically bounded cognitive capacity to explain the results of any specific trial.

[15] Consider a further worry about perception and HEC. In perception and in action based on perception, humans often think about or perceive the same things with which they interact; if the cognitive system is individuated liberally, these things are part of the subjects' minds, because the subjects' perception depends heavily on them. But this leads to a kind of idealism that should give us pause: on this view, humans do not, on the basis of perception, interact with the objects perceived; rather, certain parts of the human mind interact with other parts of the human mind!

[16] I do not claim that the primitivist view is without problems. Many authors reject Fodor's radical concept nativism because it counts too many concepts among the innate – in Samuels's

terms, it overgeneralizes (2002, pp. 256-59). Samuels responds to problems of this kind by adding the following condition to primitivism: a psychological ability, capacity, or state that is primitive yet arises only in nonstandard environments is not innate. But consider the case of cognitive deficits resulting from childhood lead poisoning. There was a time when exposure to lead threatened to become part of the developing child's normal environment – species wide – yet, even if the threat had been fully realized, having impaired cognitive functions of the relevant kind (lowered I.Q., for example) would not have been innate. And notice, this is not a far-out thought experiment; only by political action and policy changes was the problem mitigated in technologically developed countries. Even today, in many regions and cities of the world, exposure to unhealthy levels of lead is so common as to count, from a statistical standpoint, as a normal condition of childhood development.

[17] But, as argued above, such asymmetries also cut against HEC's plausibility; see note 11.

[18] Note that embedded views minimize internal representation via a different strategy than the one often employed by HEC-theorists: the extended approach lends itself to the external placement of the vehicles or bearers of representational content (Houghton 1997; Hurley 1998; Wilson, 2004), while the embedded view keeps the vehicles on the inside but, typically, binds them in a context-specific way to aspects of the immediate environment. Of course, the HEC-theorist or the advocate of the embedded approach could take an eliminativist view of representation (Brooks, 1991), more plausibly at the sub-personal than the personal level (Dennett, 1987). Note, however, that if the HEC-theorist commits to eliminativism about representation at the very level of the extended model, her talk about external vehicles is misleading (what do the vehicles bear, if not content?) (cf. Clark, 2005, note 1; Hurley 1998).

[19] Note, though, that many of these theories are designed, in the first instance, to apply to representational structures that persist and are repeatedly activated (or tokened), carrying their standard meaning on each occasion of activation – which seems to remove these theories from the running in the present context. Consider Fred Dretske's (1988) influential account of representational content. Dretske's presentation relies in a substantive way on processes that extend through time. A structure that at one time merely indicates a state of the environment acquires, by a kind of reinforcement-based learning, a role in the cognitive system – e.g., the function of controlling a particular motor response. That structure thereby becomes a representation, its content being whatever property or state the structure indicated such that the structure's indication of that property or state explains why the structure acquired its motor-controlling role in the cognitive system. Granted, although Dretske's presentation typically suggests longer time scales, the relevant learning process *could* occur very quickly. Keep in mind, however, that Dretske also intends to solve the problem of misrepresentation: later tokenings of the structure in question misrepresent when the structure's tokening is caused by something other than what it indicated when it first acquired its function in the cognitive system and in virtue of the indication of which it acquired that function. Without working through the details we cannot be sure whether Dretske's theory could be plausibly adapted to cover context-specific representations, i.e., whether his story about learning could plausibly apply on very small time-scales and do so in a way that handles the problem of misrepresentation. Instead of taking on this project, though, it seems better advised to look for a theory of content that applies in a straightforward way to context-specific representations.

[20] Gallistel (1990) and Grush (1997) offer related views.

[21] Cf. Cummins's discussion of indexicality and nested intenders (*op. cit.*, pp. 118-20).

[22] Rupert (1996, 1998/1999, 1999, 2001) develops many of the details of such a view; also see the work of Fodor (1987, 1990, 1998), Margolis (1998), and Prinz (2000), for views that are amenable in one way or another to the sort of position outlined in the text. Note that the present question concerns primitive central representations, not structured central representations that somehow copy the structure of complex sensory representations; thus, the antiempiricist objections raised by Fodor (2003) do not apply.

[23] If the human organism is innately endowed with the tendency to create or structure environments so as to lead to concept acquisition in the manner described in the text, then a further innately determined source contributes in a significant way to the content of central representations.

[24] Work on this chapter was supported by a National Endowment for the Humanities Fellowship for College Teachers. My thanks to the NEH. I would also like to thank Edward Averill and Robert Wilson for comments on an earlier draft.