Coining Terms in the Language of Thought: Innateness, Emergence, and the Lot of Cummins's Argument against the Causal Theory of Mental Content*

(Appears in *Journal of Philosophy*, October, 2001)

Robert Cummins argues that any causal theory of mental content (CT) founders on an established fact of human psychology: that theory mediates sensory detection. He concludes, "CT is false in all its forms."[1] Cummins takes his conclusion to follow from general characteristics of human sensory systems and CT; thus, he does not discuss in detail the nature of the theories that mediate sensory detection or the nature of the mental representations that, according to CT, are tokened as the result of humans' sensory interactions with the world. I contend that the details matter: when we fully appreciate the range of ways in which mental content and the *vehicles* of mental content may plausibly emerge in the human cognitive system, CT appears, after all, to be consistent with theory's mediating role in human perception.

I. Cummins's Argument against CT

Generally speaking, CT asserts that the content of a primitive mental representation, or term in a language of thought (LOT),[2] is fixed by its causal relation to kinds or their individual members (or properties or their instantiations; assume this qualification throughout). Proponents of CT have formulated a variety of complex theories of content, each theory specifying the content-fixing causal relation in its own way.[3] Given the variety of such approaches, it is not entirely clear how to properly circumscribe causal theories, or to what extent there is a substantive philosophical division waiting to be made between true causal theories and the rest. This is, however, largely a side-issue. Cummins's critique takes aim at one specific feature shared by many theories of content, a feature Cummins places at center stage when describing CT:

Cummins defines CT as "the doctrine that the contents of the semantic primitives in the human scheme of mental representation are determined by their role in detection," (535) where, in particular, the determination of content rests on the subject's ability to *reliably* detect members of the kind in question. Thus, whether we classify a given theory of content as a causal theory matters little for present purposes: Cummins's critique of CT simply takes as its target any theory for which reliable detection is a necessary condition of correct content assignments to human mental representations; the literature contains enough theories of this sort to merit careful and critical consideration of Cummins's argument.

Cummins claims that the mediating role played by theory in human perception creates an insuperable problem for CT. Currently accepted theories of human perception imply that for a subject to reliably detect members of a kind, she must possess a theory of that kind (N.B., this attribution of theory does not imply that the subject consciously holds beliefs with the content expressed by the theory). Cummins labels this condition on detection the *nontransducibility of distal properties*, or *NTDP*. A difficulty arises for CT because, in many cases, the theory required to mediate reliable detection must be explicitly represented and learned. In such cases, CT cannot explain how the subject comes into a content-fixing causal relation to the relevant kinds or their members: to construct a theory that will allow her to reliably detect Cs, the subject must possess a concept with the content *C*, so that *C* can appear in the subject's explicitly represented theory of Cs. Yet, the subject cannot possess the concept *C* in the absence of a theory that allows her to reliably detect Cs. Without the theory, there is no concept; without the concept, the subject cannot formulate the theory. Thus, CT fails as a general theory of content for human mental representations. Cummins conveys this point by way of an example: The typical human can reliably detect cats in the environment but does not do so directly: much

cognitive processing comes between sensory stimulation and the subject's tokening of a LOT term with the content *cat*. As Cummins puts it, "To detect cats...requires a theory that says, in effect, what sorts of proximal stimuli are reliable indicators of catness. To detect cats visually, you have to know how cats look." (536) Alas, the subject can have no explicit knowledge of how cats look *sans* the ability to represent cats. Thus, for many human concepts of just the sort to which CT is supposed to apply, CT does not provide a viable theory of content; together with the facts of human perceptual psychology, CT implies that you must possess a concept of cats as a means of acquiring a concept of cats, which is impossible.

II. Innate and Unlearned LOT Terms

Cummins restricts the scope of the preceding argument: if there are LOT terms whose contents are fixed independently of explicit-theory acquisition, perhaps a causal theory of content applies to such terms. Here Cummins acknowledges the possibility that some LOT terms are innate, their content fixed independently of their appearance in explicit theories (Cummins suggests 'square' as a candidate--537). Still, causal theorists should postpone the celebration: given what would appear to be the limited stock of innate terms, Cummins considers it unlikely that nativism will rescue CT (537).

It is intuitively plausible that humans possess a relatively limited range of innate concepts, yet we should wonder whether this intuitive reaction can yield the result Cummins claims for it. It is notoriously difficult to pin down the precise meaning of nativist claims,[4] which of course makes it difficult to evaluate such claims. What is more, the issue of innateness bears only indirectly on the present debate. When evaluating Cummins's argument against CT, the category of importance to the causal theorist--the category of terms that escape Cummins's attack--consists of terms that acquire content in the absence of explicit theories of those things to

which the terms apply. The force of Cummins's critique of CT rests squarely on cases where explicit theory mediates detection, where explicit theory must be present if the content of the pertinent LOT terms is to be fixed per CT. The greater the number of primitive LOT terms that can be acquired in the absence of explicit theory, the less destructive the causal theorist should find Cummins's argument. Let us use *learned terms* to refer to those LOT terms for which, assuming CT and NTDP, it is a necessary condition of their acquiring content that the human subject formulate an explicit theory of that to which the terms apply; call all other LOT terms *unlearned*. The present point can then be expressed as follows: We should not assume that the categories of unlearned terms and innate terms are coextensive.[5] Although some definitions of 'innateness' may be weak enough to apply to all unlearned terms, Cummins cannot rest happily defining 'innateness' in a permissive way. The specific purposes to which Cummins puts antinativist claims demand a strong definition of 'innateness'. When, for example, Cummins points out the obvious poverty of our innate conceptual endowment, he appeals to an intuition whose force rests on a strong notion of innateness: it is because we have in mind a demanding notion of innateness that the poverty of innate concepts seems so clear. If we limit our discussion to the more demanding definition of 'innateness' Cummins's arguments require, we have no reason to think that innate terms exhaust the contents of the class of unlearned terms. In what follows, I shall have a good deal to say about innateness and the role antinativist claims play in Cummins's critique of CT. Nevertheless, we should not lose sight of the more fundamental point, concerning the scope and composition of the category of unlearned terms.

Sections III and IV establish a general conclusion: the category of unlearned terms seems likely to extend much farther than Cummins acknowledges. These sections do not alone rebut Cummins's critique of CT, for no matter how large the category of unlearned terms, there likely

remain at least some primitive, learned LOT terms; if CT is to possess a satisfying degree of generality, the causal theorist must explain how causal relations can fix the content of primitive, learned LOT terms while avoiding the apparent circularity Cummins identifies. To this end, I argue in section V that Cummins underestimates the theoretical resources available to the causal theorist in her attempt to explain LOT term acquisition: The initial framing of theoretical axioms that facilitate detection of Cs should be understood to include only a *nonsemantically individuated* LOT term 'C'. Individuating a newly coined LOT term according to content-independent criteria allows us to see how a primitive LOT term can be introduced and its meaning fixed by theoretically mediated detection, without the term's having to possess content as a condition of its appearance in theoretical axioms

III. Innateness and the Emergence of Content

While Cummins accepts that some LOT terms may be innate, NTDP is non-negotiable: theory mediates even the causal relation between innate terms and their contents. What is the nature of such theory? Are infants born possessing a theory of squares, sufficiently detailed and accurate to facilitate the reliable detection of squares? In a footnote, Cummins remarks on mediating theories from the standpoint of the LOT theorist: "According to LOT, if it [the mediating theory] is learned, it will be a set of sentences in LOT. If it is innate, it might, in some sense be implicit in the architecture." (536, n. 5) The former case drives Cummins's negative argument, yet the latter deserves more attention. For the causal theorist wishing to lessen the force of Cummins's critique, the idea of an implicit theory seems to hold promise: if innate, implicit theory can, in the case of 'square', play the detection-mediating role required by CT and NTDP, why should not implicit theory, *innate or otherwise*, play such a role in the case of 'cat'?

Cummins briefly considers the idea that innate, implicit theory plays the requisite mediating role in the case of 'cat', but he cannot accept the implications of this view: in his words, "[W]e are not born knowing how cats look...We must, then, learn the theory that mediates cat recognition" (537). To give the discussion focus, I offer the following reconstruction of Cummins's response to the nativist defense of CT:

(P1) CT can only account for the fixation of the content of the primitive LOT term 'cat'[6] by appealing either (a) to what the human subject knows at birth about cats or (b) to the human subject's acquisition of an explicit theory of cats.

(P2) What the human subject knows at birth does not fix the content of her primitive LOT term 'cat'.

Therefore, an explicit theory of cats must be acquired, if CT provides the correct account of how the human subject acquires the primitive LOT term 'cat'.

As Cummins presents the argument, and as I have formulated it, the conclusion pertains specifically to 'cat'. Yet for the argument to fulfill its purported function, it must apply to a wide range of primitive LOT terms: Cummins intends his overall argument to constitute a decisive critique of CT; a single counterexample to CT, while providing a challenge to causal theorists, would hardly devastate their research program. Presumably, then, Cummins chooses 'cat' somewhat arbitrarily, 'cat' standing in for a large class of primitive LOT terms all of which, assuming CT and NTDP, can only acquire content via the mediating effects of explicitly

represented theory (and thus, according to Cummins, are such that CT cannot account for the fixation of their content). In light of the generality of Cummins's point, one should at least begin to doubt that (P1) canvasses all of the ways in which a causal theorist might plausibly account for the fixation of the content of a primitive term in a human's LOT. In the subsections that follow, I offer specific reasons to think that (P1) fails to cover all of the relevant possibilities, or that if it does cover them all, it does so by presuming a weak interpretation of what it is to be innate or to know something at birth--so weak that (P1) and (P2) cannot legitimately collaborate in support of Cummins's conclusion.

A. Cummins rightly allows that innate, implicit theory might play the mediating role required by the conjunction of CT and NTDP; why, one may wonder, should the content-fixing power of implicit theory be limited to *innate* implicit theory? If implicit theories that mediate reliable detection can appear over time in the postnatal subject, (P1) does not cover all cases. Let us dub as *E.I.-terms* ('E' for emergent, 'I' for implicit) all primitive LOT terms whose contents are fixed, per CT, via the mediating effects of implicit theory that is not present at birth. That the human cognitive system contains E.I.-terms seems quite plausible, given the variety of ways in which information can be implicit[7] and the variety of ways in which the human cognitive and biological systems change over time. At the very least, implicit theories could emerge as "maturational changes" themselves dependent on the organism's prior interaction with its environment (for that matter, an implicit theory could appear as the result of a subject's suffering a head injury--the effect of a certain kind of unpleasant interaction with the environment).[8]

In fairness to Cummins, we should pay careful attention to his description of implicit theory: it is "in some sense, implicit in the architecture," according to CT-cum-LOT (536). If we take

cognitive, or functional, architecture to be fixed, and if by this we mean unchanging, then no implicit theory at all can emerge; emergence is a form of change. Cummins may well have this conception of functional architecture in mind. Elsewhere in his recent writings, when Cummins talks of functional architecture, he refers to it as "fixed," citing the work of Zenon Pylyshyn. There Cummins takes Pylyshyn's view to be representative of the computationalist's perspective,[9] a perspective we can safely assume Cummins attributes to the causal-cum-LOT theorist. However, the sense in which Pylyshyn wishes to hold fixed the functional architecture will not do the duty Cummins expects of it.

Pylyshyn holds functional architecture fixed for explanatory purposes:

> [T]he architecture must form a cognitive "fixed point" so that differences in cognitive phenomena can be explained by appeal to arrangements (sequences of expressions and basic operations) among the fixed set of operations and to the basic resources provided by the architecture...If the functional architecture were to change in ways requiring a cognitive rule-governed explanation, the architecture could no longer serve as the basis for explaining how changes in rules and representations produce changes in behavior.[10]

Appeals to functional architecture are supposed to yield explanatory power by *accounting for* cognitive processes: such appeals are intended to explain semantically characterized processes in terms of series of applications of nonsemantically characterized, simple operations, which operations are not themselves affected by processes described in semantic or representational terms (using Pylyshyn's now well-known description, these operations are not "cognitively

penetrable").[11] This provides an important sense in which the functional architecture is fixed: functional architecture is not subject to changes that must be characterized in cognitive terms; instead, functional architecture's being mere physical mechanism, following laws of nature stated in nonsemantic terms, explains how cognitively characterized behavior can arise from, or be instantiated in, an underlying physical system.[12]

The preceding picture of functional architecture and its role as an explanatory fixed point does not require that each aspect of the functional architecture be either innate or unchanging. Pylyshyn describes well the way in which the architecture can change and the way in which such changes can influence higher levels, i.e., the symbolic (sometimes called 'syntactic') and ultimately the semantic levels:

> Thus biological factors interact with symbol level generalizations by causing modifications in the basic computational resources, which I call the *functional architecture*. Such modulations may be due to biochemical influences, maturation, the triggering effect of environmental releasers, dendritic arborization, atrophy of neural function from disuse, and so on.[13]

Cummins's view seems insensitive to the possibility Pylyshyn describes: the possibility of a change in functional architecture that does not result from cognitive causes. Consider the following comment of Cummins's on the representational stock and its relation to architecture:

> What Fodor has given us is an argument that primitive *representations* cannot be learned, that is, that one's basic representational power cannot be altered by

> learning. With this I am in complete agreement: One's representational repertoire is part of one's innate architecture.[14]

Cummins seems to assume that if a concept is not acquired by cognitively described means (as the result of learning, in particular), then the concept is innate. We should reject this assumption, given that (a) Cummins takes a concept's innateness to entail knowledge, present at birth, about that to which the concept applies, (b) theory implicit in the architecture can mediate content-fixation (i.e., play the role of knowledge, in "knowledge, present at birth"), and (c) architecture can change postnatally in ways that do not consist in learning but nonetheless result in the subject's holding new implicit theories.

Pylyshyn's promising explanatory strategy places a certain constraint on our characterization of the functional architecture: In order that we satisfactorily explain cognitive processes by appeal to the elements of the cognitive architecture, we should not characterize that architecture itself in cognitive terms. Cummins's view might seem to cohere well with Pylyshyn's: we cannot explain the acquisition of new primitives *as the result of learning*, for this would undermine Pylyshyn's explanatory strategy. However, Pylyshyn's view, and I would claim the computational approach in general, should allow the appearance of new "representational" primitives *as the result of noncognitive causes.*

According to Pylyshyn, a representational primitive can be identified with "a physical code or a symbol."[15] To count as part of the functional architecture, it must be individuated in the former terms: "Being instantiated in the functional architecture merely means being explainable without appeal to principles and properties at the symbolic or the semantic level."[16] While it is useful to organize a discussion of the cognitive system around different levels--biological, syntactic, semantic--the present context demands investigation of interlevel relations. In this

regard, Pylyshyn argues that changes at the physical level alter the resources available to be described at the cognitive level; primitive operations and representations can emerge in the system largely as a result of changes at the physical level, but the physical changes themselves have no relevant cognitive or semantic characterization (I develop this idea in more detail in sections IV and V, below). Cummins can, if he likes, deny the sort of interrelation Pylyshyn describes. Cummins might wish to claim that all implicit theory ever mediating the fixation of content in keeping with CT is fixed in the cognitive architecture at birth, never changing. However, given that the acceptability of (P1) depends on such a strong claim, Cummins had best offer significant considerations in its favor; he has not done so. If, on the other hand, Cummins countenances the sort of interrelation between levels that Pylyshyn describes (where new architectural features can emerge from maturation, experience-dependent synaptic growth, etc.), (P1) fails to cover the full range of content-fixing options. Thus, lacking some reason to think that all the changes in functional architecture amount to the unfolding of knowledge present at birth, human possession of E.I.-terms would seem to be a genuine possibility.[17]

B. The preceding has shown that a theory implicit in the functional architecture need not be innate in the way Cummins's argument presupposes it to be. We should also wonder why Cummins limits implicit theory to aspects of cognitive architecture. I do not suppose that Cummins intends "theory implicit in the architecture" to simply mean *whatever* in the subject facilitates reliable detection in the absence of explicit theory. But if Cummins has in mind a substantive constraint on what counts as implicit theory built into the functional architecture, he should explicitly state the criterion and argue for it. To the causal theorist any such criterion will appear suspect; to say that, relative to a given LOT term, an implicit theory facilitates content-

fixing reliable detection, need, for the causal theorist, amount to no more than saying that in the absence of an explicitly represented theory of that to which the term applies, the term nevertheless comes to stand in the content-fixing causal relation to the kind or to instances of the kind in question.[18] The causal theorist might take great interest in the mechanisms that constitute implicit theory, the operation of which contributes to the fixation of the content of unlearned, primitive LOT terms; she might also be curious to what extent these mechanisms are embodied in the cognitive architecture; regardless, she should be perfectly satisfied with my minimal characterization of implicit theory. For as regards the fixation of content, all that ultimately matters to the causal theorist is the establishment of the right causal relation. Given that CT seems to require no more than this liberal understanding of implicit theory, there seems little reason for the causal theorist to think that all content-fixing implicit theory is part of the cognitive architecture; at least Cummins has given her no reason to do so.

C. I have argued that implicit theory, construed as part of the functional architecture, need not be innate (section III, subsection A); I have also questioned Cummins's identification of implicit theory with what is built into the functional architecture (section III, subsection B). In doing so, I have tried to pry apart claims about innateness, implicit theory, and functional architecture. However, we should not neglect what would seem to be Cummins's underlying reason for discussing theory implicit in the architecture. Cummins assumes that any unlearned concept is innate and that a concept's innate status implies the presence of knowledge[19] at birth; the discussion of implicit theory, built into the architecture, then serves to explain how the requisite knowledge could be present at birth. Thus, the idea that what is innate is known at birth motivates Cummins's brief discussion of what is implicit in the functional architecture of the

cognitive system. The same commitment explains the intuitive appeal of (P1). It seems plausible that a primitive LOT term's content must be fixed either by what is innate or what is learned (I say "it seems plausible," but I do not mean to endorse this intuition); if we then accept that a concept is innate because of what is known at birth about its application, it might seem that (P1) covers all the relevant possibilities.

To dispel what intuitive appeal may still attach to (P1), I will argue that it is unfounded to equate a concept's being innate with the newborn's possession of theoretical knowledge. To this end, I now briefly survey some conceptions of innateness currently of interest to cognitive scientists, with the following question in mind: To what extent does any one of these conceptions of innateness imply that, for a subject to possess an innate concept, she must possess knowledge at birth about that to which the concept applies? I need not survey every concept of innateness ever seriously proposed. I merely intend to show that a good number of the conceptions of innateness presently on offer in cognitive science do not, taken together with CT and NTDP, sanction the inference from "*C* is innate" to "newborns possess knowledge of *C*'s application sufficient to facilitate reliable detection of that to which *C* applies." Thus, we can accommodate the common intuition that any given concept must be either innate or learned, without accepting that a concept's innateness implies the newborn's possession of the relevant kind of knowledge.

Some of the conceptions of innateness I discuss imply that *something* relevant to concept acquisition is present in the newborn, and in some cases, what is present in the newborn seems similar to belief or knowledge (see footnote 19). However, the degree of similarity falls short of what is needed to support Cummins's conclusion. Although nativist views often impute knowledge or belief to the newborn, in many cases the mental states thereby attributed differ from knowledge and belief as these are normally conceived of, to such an extent that attributions

of such states do not justify Cummins's combination of (P1) and (P2). Some special terminology will facilitate the discussion that follows: Let us call *(P2)-knowledge* the kind of knowledge Cummins has in mind when he dismisses out of hand the possibility that newborns know enough about cats to facilitate the fixation of the content of 'cat' per CT. As it turns out, a number of prominent conceptions of innateness want identification of what is innate with anything that implies the presence of (P2)-knowledge at birth.

Consider first a pair of characterizations of innateness offered by Jeffrey L. Elman, Elizabeth Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett: "When we say that a form or a behavior is innate we really mean that, given normal developmental experiences, it is a highly probable outcome."[20] Note that (1) this conception of innateness is fairly attractive, for it seems to capture the main idea behind nativist arguments from the universality, or near universality, of outcome, while (2) this definition does not require that (P2)-knowledge be present in the newborn in order for some later-emerging behavior or cognitive structure to qualify as innate;[21] on this view of innateness, nothing more than the probability that a given outcome will appear in a given environment or range of environments determines the outcome's native, or nonnative, status. As a second option--the one they officially endorse--Elman et al. propose to use 'innate' "to refer to putative aspects of brain structure, cognition or behavior that are the product of interactions internal to the organism."[22] Here again, innateness does not require (P2)-knowledge. An innate concept must emerge from organismically internal relations, but this implies no particular form of knowledge present in the newborn supporting this emergence.

Elsewhere in developmental psychology, as well as in linguistics, much has been made of what appears to be an innate basis for a variety of capacities. It has been asserted, for example,

that humans possess innate knowledge of grammar, an innate theory of mind, and an innate object concept, as well as innate domain-specific capacities for recognizing faces and distinguishing between phonemes.[23] We do not find in this collection of work a distinctive or shared definition of 'innateness'; yet given the strong nativist claims often issuing from these quarters, it would be of interest if theorists working in these areas were to make nativist claims, while avoiding, or even eschewing, the attribution of (P2)-knowledge to the newborn. I do not assert that no researcher working on these topics has ever claimed the existence of (P2)-knowledge at birth. Nevertheless, even some of the staunchest nativists hedge their claims about innateness, as is illustrated by the two following examples.

First, take the case of Noam Chomsky, probably the single most important intellectual force in the nativist revival. Chomsky intentionally distances his nativist theses from what I am calling (P2)-knowledge. While he claims that our knowledge of universal grammar is innate--it is "somehow represented in the genotype"--he is careful to remind us that 'knowledge', as he uses the term, does not necessarily imply knowledge as it is normally conceived of; to avoid confusion, he says that we "cognize" universal grammar.[24]

Second, consider the case of the child's theory of mind. While the theory of mind is frequently claimed to be innate, it is widely recognized that it takes three to four years for the theory to get up and running. Very little, if any, of the theory is present at birth (the tendency to imitate, perhaps?). It seems to make perfect sense to many researchers working in this area to take a nativist position regarding the child's theory of mind without claiming that the newborn has (P2)-knowledge of the theory.[25]

Two views of innateness remain, both of some stature. The first is intended to capture the idea that genes determine innate traits, behaviors, or concepts. Richard Samuels expresses the

view as follows: "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."[26] Without much comment it should seem clear that this characterization of innateness is perfectly consistent with, and in fact seems specifically constructed to allow, the epigenetic emergence of innate characteristics: it requires nothing in the way of knowledge, and certainly not (P2)-knowledge, in the newborn in order that a trait, behavior, or concept be innate.

The final view to be considered might well be, of the forms of nativism discussed here, closest to what Cummins had in mind when formulating his critique of CT. According to this view, a concept is innate if and only if it is primitive, where being primitive now means *not being acquired by hypothesis testing and confirmation*.[27] Cummins's discussion of innateness contrasts those LOT terms relative to which explicitly represented theory mediates content-fixation and LOT terms relative to which it does not. By identifying innate terms with those in the latter class, Cummins seems to be following Fodor; for one might think there is little, if any, distance between LOT terms not acquired by the framing of explicit theory and those not acquired by hypothesis confirmation. However, if I have accurately captured Cummins's thinking, he errs in his treatment of Fodorean primitives: Cummins gives no argument to show that a LOT term's being unlearned, in Fodor's sense of not being acquired by hypothesis confirmation, requires that the newborn have (P2)-knowledge about that to which the term applies.

In fairness to Cummins, we should note the misleading nature of some of Fodor's own remarks on the acquisition of primitives. In *The Language of Thought*, Fodor says, "My view is that you can't learn a language unless you already *know* one,"[28] going on to characterize the known language as innate. This suggests the unnecessarily strong notion of innateness assumed

by Cummins, according to which a concept's being innate implies the presence of certain bits of (P2)-knowledge at birth. Nevertheless, as Fodor proceeds, he explicitly repudiates this strong notion of innateness: "The view presently being proposed doesn't require that the innate conceptual system must literally be present 'at birth', only that it not be learned."[29]

The preceding survey, while necessarily brief, illustrates the genuine difficulty Cummins faces. His argument against CT succeeds partly by taking an expansive view of the class of learned terms and painting as hopelessly inadequate the set of primitive, unlearned terms: the latter set, Cummins implies, consists only of innate terms, about whose applications we must, at birth, possess knowledge sufficient to mediate content-fixation; and clearly, we have very few of these. However, given that contemporary cognitive science embraces a wide range of conceptions of nativism, many of which do not require (P2)-knowledge at birth, Cummins's approach seems misleading. Even if we identify the unlearned with the innate--in keeping with the intuition that any given concept must be either innate or learned--innateness does not imply the presence of the relevant (P2)-knowledge.

IV. The Emergence of LOT Terms Characterized Nonsemantically

Now consider the genesis of the vehicles of mental content. As Cummins describes LOT, "It has a finite number of semantically primitive expressions *individuated syntactically*" (535, emphasis added). I will construe the requirement of syntactic individuation as a requirement that LOT terms be individuated nonsemantically (there is, however, much more to be said about this issue).[30] The question then naturally arises, "What are the nonsemantic individuation criteria?" Although alternative approaches exist, those inclined toward realism regarding mental representations often assume that we can, in principle, state these criteria in the language of neuroscience.[31] [32] When we think of LOT terms in this way, we see that LOT terms, even those

typically classified as innate, emerge in the child's cognitive system over time: the very neural structures that are to become vehicles for content appear largely as a result of the child's interaction with her environment.

A summary excursion into developmental neurobiology may help to clarify my point. Developing brains grow and change in numerous and varied ways. Despite what looks to be a messy, nonlinear process of growth and change, stable neural structure emerges, which subserves the subject's representational capacities. In an attempt to understand this process of emergence, neuroscientists ask such questions as (1) How does the overall functional structure of the brain appear without there being a tiny map of the brain's ultimate layout contained in the genetic material? and (2) How does experience contribute to the development of the functional properties of various neural structures? A commonly given answer to the first question leads immediately to prosecution of the second. It is widely thought that overall structure emerges via the delicate interplay of multiple factors: there is, it is said, an epigenetic dance, experience appearing as one of the key participants. To a significant extent, the subject's experience shapes her neural substrate, determining the way in which the physical material of her brain eventually realizes its function as the center of cognitive processing. The interesting debate here is not over whether experience affects neural development--the experimental evidence seems to have settled that issue--but over the precise role of experience in shaping the structure of the brain.[33]

On one side in this debate stands the selectionist approach, a view closely associated with the work of Gerald Edelman.[34] According to selectionist views, genetically programmed growth provides a wealth of raw material that experience, among other factors, then winnows to create functional neural structure. Early development consists partly of the spreading of billions of dendritic and axonal arbors. Rich in synapses, these overlapping branches provide an enormous

number, in the quadrillions, at least, of possible ways in which groups of neurons might excite or inhibit each other. Of these various patterns of connections, experience reinforces and strengthens those that have proven most useful to the organism, or which have simply been used more than competing ones. The activity of strengthened patterns of connectivity overshadows the less useful or less used patterns of connections, resulting in ineffective activity in the overwhelmed connections; dendritic or axonal retraction or withering; or even cell death.

Constructivism offers an alternative.[35] Constructivists agree with selectionists that experience contributes to regressive events. However, they assign a further role to experience: it guides the generation of axonal and dendritic arbors and can selectively increase the number of synapses in a given area of the brain or a given neural circuit; experience does so at least as early as the perinatal stage and as late as adulthood. Constructivists claim that such environmental instruction causes the growth of the very neural resources that are needed to represent aspects of problems the subject will solve using those resources: the specific character of experience catalyzes the growth of the neural resources needed to more effectively process the input that triggered the development of those resources. While some selectionists accept that there are multiple waves of growth and pruning within the life span of one organism, this is the fundamental point of disagreement between the two camps: whether experience can, as constructivists claim, cause axonal and dendritic growth or increases in the number of synapses in a way that is sensitive to the content or structure of the input itself.[36]

The preceding discussion of perspectives on neural development gives some indication of the prevalence of epigenetic views among contemporary neuroscientists.[37] This points serves three purposes: (1) to underscore the false nature of Cummins's dilemma offered in (P1), by describing some mechanisms likely to govern the emergence of nonsemantically individuated

LOT terms in humans, mechanisms that would seem operative in cases of at least some primitive, unlearned LOT terms, (2) to illuminate the integrative vision above attributed to Pylyshyn, involving the purely physical emergence of new elements of cognitive architecture, which elements then subserve cognitive processes conceived of semantically, and (3) to provide background for the discussion of the following section, wherein I put to further use the distinction between LOT terms typed by their content and LOT terms conceived of nonsemantically.

Regarding any given primitive LOT term, Cummins asserts that the causal theorist must choose between its being (a) innate, the fixation of the term's content mediated by knowledge possessed at birth, and (b) acquired at least partly by the subject's formulating an explicit theory of that to which the term applies. If selectionists *or* constructivists are correct, the vast majority, perhaps all, of an adult's nonsemantically individuated LOT terms appear in no substantial form at birth. This list of nonsemantically individuated concepts not present at birth seems quite likely to include such LOT terms as 'physical object', 'subject' (in the grammatical sense), and 'belief', terms that would seem primitive and unlearned and, when categorized according to their content, are often considered innate. Many such concepts *are* plausibly innate, in some sense: they emerge in all standard subjects across a wide variety of contexts. Nonetheless, they do not fit Cummins's description of unlearned terms. They do not seem to be associated with content-fixing knowledge present at birth, even implicitly represented knowledge. How could they be? These concepts, nonsemantically individuated, have yet to appear on the developing subject's cognitive scene.[38] Thus, such terms do not seem to fall into either of the categories Cummins recognizes. The causal theorist should feel free to account for the fixation of these terms' contents by appeal to something other than the processes allowed by (P1).[39]

V. LOT Term Introduction and the Content of Learned Terms

For a primitive concept to be unlearned, the newborn need not possess a content-determining theory of that to which the term applies; furthermore, she need not possess the nonsemantically individuated LOT term that will one day carry the content by reference to which we would normally identify that concept. Or so I have argued.[40] Nonetheless, what has been said so far does not show that any particular LOT term, 'cat', for example, is unlearned; nor does it show that all LOT primitives are unlearned. Consequently, to round out my defense of CT, I must now address the difficult case, where explicitly represented theory seems necessary to mediate content-fixing reliable detection.

Assume, then, plausibly enough, that the explicit formulation of theory mediates reliable detection for 'cat', via such theoretical axioms as Cummins lists: "(A) Cats have whiskers. (B) Cats have four legs. (C) Cats have fur" (537). Is there a way for the subject to frame such explicitly represented axioms without requiring--absurdly, if we accept CT--that the subject already possess a concept with the content *cat*?

This is possible, if we envision the initial theoretical axioms as containing the nonsemantically individuated term 'cat', as opposed to a LOT term with the content *cat*. In section IV, I described some of the ways in which LOT terms, conceived of nonsemantically, might be introduced at the neural level: early human development comprises an ongoing process of neural generation, growth, and death and synaptic generation, strengthening, and weakening that results in the emergence over time of functionally coherent cell assemblies. If this picture is correct, then at least some of the time after birth, the human brain/cognitive system generates, or 'coins', new LOT terms, individuated nonsemantically: it generates new circuits or cell

assemblies of the same sort as those we describe when we characterize LOT terms in the language of the neurosciences. It is possible, then, that the subject's early theory of catness include a newly coined LOT term lacking the content *cat*, perhaps lacking content altogether. If this picture is correct, then the initial theoretical axioms that include the nonsemantically individuated LOT term we call 'cat' would more properly be thought of in the following way: (A) *t* (a newly coined, nonsemantically individuated LOT term) have whiskers; (B) *t* have four legs; (C) *t* have fur.[41] [42]

Cummins's fundamental objection to CT is thus dissolved: Take learned concept *C*. Given CT and NTDP, the subject cannot possess a concept with the content *C* unless she possesses a nonsemantically individuated vehicle 'C'; the vehicle must be present to hold a place in the subject's theory of Cs. However, 'C', characterized nonsemantically, does not have to have content *C* when it first appears in the axioms constituting the theory that mediates the fixing of content *C* to vehicle 'C'. The alleged circularity of CT vanishes.

To flesh out and defend this proposal, I respond below to a series of critical questions:

Q. #1: What causes the brain/cognitive system to produce a new LOT term that will be the subject of an explicit theory?

A. #1: When the subject encounters a cluster of properties not sufficiently similar to any stored cluster (each cluster embodying a mediating theory), her cognitive system introduces a new primitive term. The term-introduction rule could itself be either explicitly represented or implicit in the architecture (although in what follows, I assume what I take to be the more plausible possibility: that the rule is implicit in the system's workings). Of course, the system must already be capable of representing the properties in the cluster, but this does *not* imply that said cluster defines the newly introduced term (see note 42).

Q. #2: When a new LOT term is introduced, why is content not fixed immediately and permanently?

A. #2: Whether content is fixed immediately depends on (1) the mediating theory of that to which the newly introduced LOT term applies and (2) the exact nature of the causal relation supposed to underlie reference. Even if content is fixed straightway, the LOT term in question, individuated nonsemantically, is not permanently saddled with that content. Imagine that the subject encounters a cluster of properties sufficiently unlike those encountered before, and in response, the cognitive system frames theoretical axioms that include a newly coined LOT term *t*. Assume these axioms to play a sufficient mediating role that, according to a version of CT we have adopted, *t* acquires content. The subject may later interact with a new item differing substantially from the item that caused the original coining of *t*, *without the new item's differing enough from the original to cause the coining of a new LOT term*. The cognitive system responds by altering existing theoretical axioms or adding new axioms including *t*. Whether this refinement changes the content of *t* depends on the causal relation established by this refinement, and this depends on the details of the case.

Note that if a new experience causes a change in theoretical axioms in which *t* appears, this does not imply that the cognitive system explicitly represented *error* in its earlier theory of *t*. Whatever theory, implicit or otherwise, governs the introduction of new LOT terms may do so by measuring the match between a newly encountered item and stored, associated clusters of properties (such clusters are sometimes called 'prototypes'). In a case where the mismatch between the new item and stored clusters is not so great as to cause the coining of a new LOT term, the result is a value-neutral integration of the profile of the newly encountered item and whatever existing prototype most closely matches the new item's profile. *Pace* Cummins (538),

then, it is conceivable that the cognitive system alter a prototype, which embodies a theory, without the cognitive system's having to represent the prototype, prior to its alteration, as normatively mistaken. Implicit in the workings of the system is a theory that causes the coining of a new LOT term under some circumstances and the integration of new experiences and existing prototypes under others.[43] Admittedly, such an implicit theory has peculiar status, perhaps that of metatheory. Better still, we may think of the theory as part of the system's program, emerging from the arrangement of the physical matter that constitutes the cognitive system; in this case, the implicit "theory" may not be found among the cognitive system's mental states as we would normally describe them; for this reason, it may be misleading to call it a 'theory'.[44]

Q. #3: The model of concept acquisition I have suggested seems to allow for computation without representation: newly coined LOT terms would seem to be in the domain of some computational operations, despite the possibility that they represent nothing; does not the orthodox computational approach to explaining cognition proscribe such models?[45]

A. #3: Three points in response: First, it seems empirically unproductive to insist that the following holds for *any* physical state with an assigned value that falls in the domain of *any* formal operation in a computational model of *any* aspect of cognition: that physical state must itself have representational content. Computational models of stereopsis, for example, explain how the brain might construct a coherent representation of the immediate environment from two separate two-dimensional arrays, one originating at each retina. Such models do so by, among other things, assigning values to physical states and placing these values in the domain of the representation-constructing algorithms specified by the relevant computational theory of vision.[46] However, there is no reason to think that every such physical state is itself a representation;

granted, each such state is assigned a numerical value, but this does not imply that the physical state *represents* that numerical value. Whether we assign representational status to such physical states depends on the explanatory value of doing so, and sometimes there seems to be little value in such making such assignments.[47]

Second, I do not hypothesize the existence of a computational system containing no representations: in the kind of system I have been discussing, representations abound; nevertheless, some newly coined LOT terms might not be among them. Note that many of the advocates of content-laden interpretations of computational processes rest their advocacy largely on the empirical fruitfulness of semantic talk. Such theorists typically agree that computational processes *can be* characterized in nonsemantic terms. It is not incoherent to talk about computational processes in, say, syntactic terms, such theorists will admit; they argue that semantic interpretation is indispensable if our cognitive models are to fully explain human behavior or cognitive processing.[48] The present view concurs to a great extent: I do not contend the claim that many, perhaps the majority of, elements participating in the computational processes underlying human cognition possess semantic values. However, turnabout is fair play, and if the serving of empirical virtues sometimes favors nonsemantic descriptions (or as I have suggested, "mixed" descriptions), then we should accept such descriptions. In particular, if we best explain the emergence of representational primitives by hypothesizing the presence of axioms, subject to computational manipulation, containing both representing and nonrepresenting elements, then this approach should be acceptable even to those, including myself, who think that content-based characterizations of some processes are required if we are to fully understand those processes.

Here it is also important to bear in mind that newly coined LOT terms are, presumably, of the sort of structure whose instances often have content in the computationally conceived-of human cognitive system. On this view, all states that participate in cognitively relevant computational processes share certain non-content-determining, nonsemantically characterized traits; newly coined LOT terms, perhaps devoid of content, exhibit these traits and, as such, are candidates for participation in computational processes. Recalling that computation can be described at a nonsemantic level, we should not automatically exclude these structures from participation in computational processing; instead we should see them as potential representations, which have not yet, but may eventually, come into content-conferring relations to kinds or properties. It is in this sense that newly coined LOT terms can be protorepresentations, appearing in axioms upon which computational processes can operate, even when the content of the incipient representations is not yet fixed. These protorepresentations may bear the physical marks of computed-upon representations before such protorepresentations fully come into the causal relation that will, according to CT, fix their content.[49]

Lastly, allowing *some* computation without representation does not imply an unacceptable proliferation of cognition; the conditions for being a cognitive system are much more stringent than simply that a system sometimes compute something. "No computation without representation" is a fine slogan when understood properly: nothing described as a computational system counts as a cognitive system unless it exhibits the kind of behavior or capacities best explained by positing internal representations over which at least some of the computations operate. So far as can be presently told, few, if any, nonhuman systems meet this condition.[50]

VI. Conclusion

If CT is to maintain any plausibility as a theory of content for human concepts, it must respect the known facts of human psychology, including human perceptual psychology. However, Cummins errs in thinking that the demand for psychological realism supports his case against CT. To show this, I first considered the ways in which concepts might emerge in the cognitive system. Cummins sees only two options: for any given concept, the causal-cum-LOT theorist must choose between saying (a) the subject has knowledge, at birth, of that to which the concept applies and (b) the concept's content is fixed via the mediating effects of an acquired, explicitly represented theory of that to which the concept applies. I have argued that this range of options is artificially limited: we should recognize the wide range of cases in which mental content and the vehicles of mental content emerge after birth in the absence of explicit theory. Cummins admits the power of implicit theory to mediate content-fixing reliable detection in the case of concepts about whose applications we possess knowledge at birth, yet it appears that implicit theory's detection-mediating power extends well beyond these cases. Correlatively, we should doubt that explicitly represented theory participates in content-fixation in as many cases as it has often been thought to. This point weakens Cummins's argument against CT, for the argument's force rests, to a significant extent, on the assumption that explicit-theory formation is a necessary condition of content-fixation for a large class of human concepts.

More damning to Cummins's case against CT was the argument of the penultimate section: Even where explicit theory contributes to the fixation of content, there exists a possible route to content-fixation within the framework assumed by CT, a route that Cummins fails to consider. It is agreed on all sides that the LOT theorist must individuate LOT terms nonsemantically; plausible ways of doing so allow us to see how a LOT term can appear, without content, in the

framing of a theoretical axiom. This possibility defeats Cummins's claim that where explicitly represented theory mediates reliable detection, CT is trapped in an inescapable circle. For the time being, then, the causal theory of mental content stands as a live option.[51]

Robert D. Rupert

Texas Tech University

---

[*] I would like to thank Edward Averill, Robert Cummins, J. Robert Loftis, and this journal's anonymous referee for helpful comments on an earlier draft of the present paper. I also thank, for their constructive feedback, audiences at Texas Tech University and the Northwest Conference on Philosophy (1999), to whom I presented earlier versions of this work. I would like to acknowledge particularly the contributions of Frederick Suppe, Edward Averill, Aaron Meskin, J. Robert Loftis, Mark Webb, and Dan Kaufman (Texas Tech) and Todd Jones, David Beisecker, Hollibert Phillips, David DeMoss, and Camille Coffey (Northwest Conference). In my early thinking about nativism, I benefited greatly from conversations with Charles Chastain, especially our discussions of Fodor's nativist views.

[1] "The LOT of the Causal Theory of Mental Content," *Journal of Philosophy* 94 (1997): 535-42, p. 541. All page references are to this work, except where otherwise indicated.

[2] See Jerry Fodor, *The Language of Thought* (Cambridge: Harvard University Press, 1975). Sometimes, when it seems convenient for expository purposes, I refer to mental representations as 'concepts', without thereby intending to take any position regarding the nature or existence of concepts as abstract, as opposed to psychological, entities.

Whether LOT exists is a contentious issue. Cummins argues that CT entails at least a minimal version of the language of thought hypothesis (539-41). While his argument merits attention, I set it aside here. I do not defend CT by attempting to sever it from LOT, as causal theorists might propose to do in response to Cummins's critique of CT; consequently, for present purposes, I grant LOT's existence.

[3] Fred Dretske and Jerry Fodor have developed two of the most prominent causal theories; see Dretske's *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981) and *Explaining Behavior* (Cambridge: MIT Press, 1987); and Fodor's *Psychosemantics* (Cambridge: MIT Press, 1987), Chapter 4, and *A Theory of Content and Other Essays* (Cambridge: MIT Press, 1991), Chapter 4.

[4] See Part I of Fiona Cowie's *What's Within: Nativism Reconsidered* (New York: Oxford University Press, 1999) for an extended discussion of this problem; also see A. Karmiloff-Smith, K. Plunkett, M. H. Johnson, J. Elman, and

E. A. Bates, "What Does It Mean to Claim that Something Is 'Innate'? Response to Clark, Harris, Lightfoot and Samuels," *Mind and Language* 13 (1998): 588-97.

[5] Arthur Danto makes a similar point in his "Semantical Vehicles, Understanding, and Innate Ideas, "in S. Hook (ed.), *Language and Philosophy* (New York: New York University Press, 1969), pp. 122-37. "Obviously, the move from *unlearned* to *innate* is abrupt, inasmuch as learning is but one avenue of acquisition..." (ibid., p. 128).

[6] Following Cummins, I set aside the possibility that 'cat' is compound. Cummins rightly sees it as very unlikely that other LOT primitives define 'cat' (537).

[7] See Cummins's "Inexplicit Information," in M. Brand and R. M. Harnish (eds.), *The Representation of Knowledge and Belief*, (Tucson, Arizona: University of Arizona Press, 1986), pp. 116-26, especially pp. 119-20. Of special interest here is Cummins's discussion of domain implicit information, which seems especially likely to emerge over time in the subject. Here one should also take into account much recent work on embedded cognition. This work encourages a minimalist view of the amount of internal, explicit representation necessary for a cognitive system to exhibit a full range of beliefs or possess a standard complement of knowledge or information. Those who view cognition as embedded claim that, in a surprising number of cases, we best explain human behavior and cognitive capacities without appealing to a rich set of internal representations and complex internal computational processes manipulating those representations; although some internal representations are admitted in these cases, it is claimed that the standard subject's cognitive states and capacities are, to a great extent, "off-loaded," i.e., stored in or heavily dependent upon the environment, and thus emerge from the relations the subject bears to her environment and its structure; if implicit theory is embodied in the relations themselves, it is easy to see how implicit theory could emerge over time, as the subject *comes to bear* the relevant relations to her environment. On embedded cognition, see Andy Clark, *Being There: Putting Brain, Body, and World Together Again* (Cambridge: MIT Press, 1997); Ron McClamrock, *Existential Cognition: Computational Minds in the World* (Chicago: University of Chicago Press, 1995); David Houghton, "Mental Content and External Representations," *Philosophical Quarterly* 47 (1997): 159-77; A. Clark and David Chalmers, "The Extended Mind," *Analysis* 58 (1998): 7-19; Mark Rowlands, *The Body in Mind: Understanding Cognitive Processes* (Cambridge: Cambridge University Press, 1999). Some of this work argues for the stronger thesis that cognition is extended: states and processes of the human cognitive system literally comprise parts of the world beyond the boundary of the organism. While more tendentious than the claim that human cognition is embedded, the view that cognition is extended also supports the present point.

[8] Cf. Fodor, "The Present Status of the Innateness Controversy," in *Representations* (Cambridge: MIT Press, 1981), pp. 257-316, p. 275.

[9] *Representations, Targets, and Attitudes* (Cambridge: MIT Press, 1996), p. 31 (*RTA*, hereafter). Also note Cummins's citation of Pylyshyn's work when discussing functional architecture and transduction (536, note 4).

[10] Zenon Pylyshyn, *Computation and Cognition: Toward a Foundation for Cognitive Science* (Cambridge: MIT Press, 1984), p. 114; also see p. 131.

[11] Of course, this explanatory approach is not the only one possible. For some purposes, it may suffice to classify as basic certain semantically characterized processes and representations, explaining all other cognitive capacities by appeal to operations and arrangements of the basic, semantically characterized ones. However, this approach does not allow an account of the workings of the entire cognitive system in noncognitive, nonsemantic terms.

[12] Among its other virtues, Pylyshyn takes this explanatory gambit to hold promise of at least a token reduction of cognitive processes; ibid., pp. 132-34.

[13] Ibid., p. 259. See also p. 260, where Pylyshyn identifies control structures as part of the functional architecture. In "Inexplicit Information," Cummins argues that information can implicitly reside in control structure (and, presumably, enough such information could constitute implicit theory). If functional architecture can change over time, and control structure is part of the functional architecture, and control structure can instantiate implicit theory, then implicit theory need not be innate in the sense of being present at birth.

[14] *RTA*, p. 145.

[15] *Computation*, p. 26. For two prominent computationalists' expression of a similar view, see Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," in J. Haugeland, *Mind Design II: Philosophy, Psychology, and Artificial Intelligence* (Cambridge: MIT Press, 1997), pp. 81-110, p. 85 (reprinted from the *Communication of the Association for Computing Machinery*, 19 [March 1976]: 113-26).

In the passages quoted here, Newell and Simon--and Pylyshyn--speak specifically of the nature of explicit representational primitives, and one might wonder about the connection between the postnatal emergence of such representations and our present concern: changes in functional architecture that result in the emergence of new implicit theories. The relation between the two issues is twofold: First, note our position in the dialectic; at this point, I intend to establish the possibility that functional architecture can change postnatally. To do so, it will suffice to have established any form of postnatal change in functional architecture, including as a possibility, the emergence

of primitive explicit representations. Second, we should also consider the possibility that, at least in some cases, it is at least partly via this very kind of change, i.e., the emergence of primitive representations, that a new theory implicit in the architecture can come to be held by the subject.

[16] *Computation*, p. 132.

[17] One could claim that all such architectural changes do, in fact, amount to the unfolding of knowledge present at birth; however, in order to legitimate such a claim, one would have to offer a supporting view of what it is for knowledge to be present. The following view--echoing one of the nativist targets of Book I of Locke's *Essay* (see Chapter 2, §5)--would suffice: *Knowledge of As is present in a subject if and only if either(a) the subject explicitly represents knowledge of As, (b) the subject is so constituted that, given certain experiences, she will someday acquire a concept with the content A that will appear in an explicit representation of knowledge about As, or (c) the subject is so constituted that, given certain experiences, she will someday acquire implicit knowledge of As.* This view justifies the claim that all primitive, unlearned LOT terms refer to things about which subjects know something at birth, notwithstanding postnatal changes in architecture, yet it does so at great cost: we know something at birth about all things--including cats--some knowledge of which we will ever explicitly represent. Cummins must reject this strategy; this cannot be the sort of knowing Cummins has in mind when he treats the claim that "[W]e are not born knowing how cats look" as standing in need of no support. Cummins might have some other, general view in mind that would legitimate the claim that all physical changes in functional architecture amount to the unfolding of knowledge present at birth, but I can not make it out; it is difficult to see how any view could be permissive enough to support the claim all architectural changes amount to the unfolding of knowledge present at birth, while being at the same time restrictive enough to support Cummins's antinativist views regarding our knowledge of that to which such concepts as 'cat' apply.

[18] The causal theorist might, with some justification, approach these matters differently: She might acknowledge substantive constraints on what is to count as implicit theory, but also argue that NTDP is false, that there are cases where we gain nothing by characterizing in semantic terms (as, say, bits of theory implicit in the functional architecture) whatever mediating machinery facilitates reliable detection. (This point was suggested by a comment of Aaron Meskin's.) I have, however, chosen to accept NTDP for the sake of argument and to criticize Cummins's position on other grounds.

[19] For an important qualification see 536 (also, *RTA*, p. 132). Here Cummins acknowledges his use of 'know' to be unorthodox in some respects, for it is the cognitive scientist's use, rather than the philosopher's. Whereas the philosopher's use implies the truth of what is known and also that it is justified, to know *P* in the cognitive scientist's sense, the subject need only treat *P* as if it were reliable information (i.e., the subject need only believe *P*, in the not very demanding sense of 'believe' that allows you to believe that *P* without any conscious access to the content of *P* and without your being able to employ the belief that *P* in a wide range of contexts). For present purposes, Cummins's semantic clarification is moot. To be true to traditional philosophical usage, some may wish to substitute 'believe' (or given the lack of any requirement of conscious access, 'believe*'; or, following Chomsky, 'cognize') in place of 'know' when we encounter the latter in Cummins's argument. Then, however, I will simply translate the argument in the text into talk about believing (or believing*, or cognizing), without any loss of force: We should still want to know why being unlearned and primitive implies the possession at birth of certain beliefs (or beliefs* or cognizings). Equivocation continues to dash Cummins's argument: Only if Cummins has in mind a standard meaning of 'belief' does the altered (P2) seem as obviously correct as Cummins takes it to be, yet being primitive and unlearned only seems to stand any chance of implying the presence, at birth, of beliefs that mediate reliable detection, in keeping with (P1), if 'belief' is used in a nonstandard, permissive way.

[20] *Rethinking Innateness: A Connectionist Perspective on Development* (Cambridge, MA: MIT Press, 1996), p. 319; see also p. *xiii*.

[21] It is a problem that discussions of innate knowledge often proceed absent any clear measure of what it is for knowledge to be explicitly represented, at birth or otherwise, and any precise view regarding to what extent and in what ways implicit representation can ground legitimate attributions of knowledge (or belief). Lacking such criteria, it becomes difficult to evaluate a claim that some bit of knowledge is or is not present at birth; see Andy Clark, "What's Knowledge Anyway?," *Mind and Language* 13 (1998): 571-75. Clearly, however, even if we count as knowledge the presence of whatever constraints lead to highly probably outcomes, this need not be (P2)-knowledge.

[22] Op cit., p. 23. The definition I initially quoted characterizes what Elman et al. sometimes call 'primal' behaviors or cognitive forms.

[23] There is a large literature addressing these issues. For starters, see *Cognitive Science* 14 (1990), which includes a series of papers by prominent developmental researchers investigating domain-specific theories and constraints, followed by Frank Keil's insightful metacommentary; also see various essays in S. Carey and R. Gelman (ed.) *The*

Epigenesis of Mind (Hillsdale, NJ: Lawrence Erlbaum Associates, 1991); Elman et al. (op. cit., pp. 107-8) list many of the domains in which innate knowledge or constraints have been thought to be operative and give references to much of the research taken to support nativist claims in these domains.

[24] See, for example, *Rules and Representations* (New York: Columbia University Press, 1980), especially pp. 69-70, 90-92; the passage quoted in the text appears on p. 91. Here Chomsky expresses an ambivalent attitude toward knowledge as it is normally conceived of and its relation to cognizing. Compare: "In fact, I don't think that 'cognize' is very far from 'know' where the latter term is moderately clear..." (ibid., p. 70), and "In fact, it is not at all clear that the ordinary concept of 'knowledge' is even coherent..." (ibid., p. 92) (I do not mean to suggest that Chomsky's views are inconsistent; perhaps the ordinary concept of knowledge is incoherent overall, although there are certain limited uses of "knowledge" that are "moderately clear.") Chomsky makes little effort to resolve his ambivalence because, as he sees matters, it is of no import to his nativist position whether cognizing and knowing (as it is normally conceived of) are the same mental state type, or whether any sense can ultimately be made of the common conception of knowledge (ibid., p. 92). For present purposes the important point is this: Chomsky does not think that a plausible nativist position in linguistics must imply the presence at birth of (P2)-knowledge of universal grammar.

[25] Consider Stephen Stich and Shaun Nichols's revealing remark about our innate knowledge of specific content domains, including our theory of mind: "[I]t is plausible to suppose that natural selection has provided the child with lots of help--either in the form of innate knowledge structures or in the form of special purpose learning mechanisms." Stephen P. Stich, *Deconstructing the Mind* (New York: Oxford UP, 1996), p. 151; this disjunctive offering suggests the extent to which the claim of an innate theory of mind is neutral in respect to the amount or kind of knowledge present at birth. For an overview of research on the child's theory of mind, see Janet Wilde Astington, *The Child's Discovery of the Mind* (Cambridge: Harvard UP, 1993); at pp. 166-67, Astington notes the gap between nativism about the theory of mind and the stronger claim that the newborn *knows* the theory of mind.

[26] Richard Samuels, "What Brains Won't Tell Us about the Mind: A Critique of the Neurobiological Argument against Representational Nativism," *Mind and Language* 13 (1998): 548-70, p. 564.

[27] Jerry Fodor has been known to frame the debate between nativists and empiricists as a disagreement about the number of primitive concepts, taking primitive concepts to be those not learned by hypothesis confirmation (in effect, those not reducible by definition to others); nativists claim there are lots, while empiricists claim that such

concepts constitute a much smaller portion of our conceptual repertoire (the sensory concepts only); see "The Present Status...," pp. 281-83, 315.

[28] *The Language of Thought*, p. 65.

[29] *The Language of Thought*, p. 96. Fodor sometimes separates these issues to an even greater extent, so that a concept's being acquired by a process other than hypothesis testing, by brute-causal triggering, for example, does not itself imply the concept's innateness; see *Concepts: Where Cognitive Science Went Wrong* (Oxford: Oxford/Clarendon Press, 1998), p. 137; also see Cowie, *What's Within*?, p. 104ff. This is not merely a point about the Fodor's intellectual development. There is good reason not to infer a concept's innateness solely from the fact that the concept is acquired via a brute-causal process: Consider Samuels's example of learning Latin by taking a Latin pill, which creates mastery of Latin in the consumer via a brute-causal process; just because Latin is thereby acquired hardly implies that knowledge of Latin was innate (or that the subject knew anything about Latin at birth) (op. cit., p. 554, n.6). (This would appear to be a variation on Arthur Danto's Spanish-pill example, which Danto uses to support the point that being unlearned does not imply being innate; op. cit., p. 128.) Note, though, that Fodor does not necessarily feel the weight of the preceding point; his reason not to infer innateness from brute-causal acquisition seems less robust, following more from considerations of what would be "a good thing tactically" (see "Doing Without What's Within: Fiona Cowie's Critique of Nativism," *Mind* 99 (2001): 99-148, p. 131).

[30] Some of which I will say here. In characterizing LOT terms, we do best to observe the two following requirements, which neither singly nor jointly imply that *syntactic properties alone* serve to individuate LOT terms nonsemantically: (1) In order that LOT be sufficiently language-like to be called a 'language' (perhaps, for example, so that it manifests combinatorial rules supporting productivity), atomic units of LOT must be assigned to categories roughly analogous to the syntactic categories of natural language, and (2) for LOT to remain true to the quasi-reductive, naturalistic project it is normally taken to be part of, there must be *some* way of individuating LOT terms that does not depend on the terms' semantic contents. Both constraints may be satisfied without its being the case that LOT terms are individuated solely by syntax (understanding syntax narrowly, to comprise only functional relationships holding among LOT terms). Say that LOT terms are individuated by reference to something other than, or more than, their syntactic roles, for example, by reference to neural types. Condition (2) would clearly be met. Also, in keeping with condition (1), each LOT term may well be of (at least) one determinate syntactic type (e.g., noun, verb--or their LOT analogues), yet a term's having this, as well as other syntactic properties, need not

constitute the term's identity conditions. For discussion of syntactic individuation and related issues, see Michael Devitt, "A Narrow Representational Theory of the Mind," in W. G. Lycan (ed.), *Mind and Cognition: A Reader* (Oxford: Blackwell, 1990), pp. 371-98.

To pursue the argument of the present section, I need only assume the satisfaction of the weak condition that there should exist nonsemantic, although not necessarily syntactic, individuation criteria for LOT terms: when the LOT theorist says, "*That item* has content *C*," I expect the LOT theorist to be able to describe what *that item* is, in some way other than by reference to *C*. Note that this condition does not demand anything nearly as strong as might be demanded by a species-wide type-type identity criterion, which is often construed to require that a given LOT term individuated nonsemantically should be found in all the various members of the same species and found to have the same content in all members of the species. In order that quasi-reductive cognitive theories invoking LOT possess the explanatory value they are supposed to, these theories would seem to have to assume no more than a certain amount of intrasubjective stability in the relationship between the subject's LOT terms individuated nonsemantically and the contents of those terms: A satisfactory explanation of the way in which an individual subject's cognitive system implements a given content-based psychological generalization might appeal only to the repeated appearance of a vehicle possessing the same content *in that subject* on different occasions when the subject instantiates the regularity in question (the precise degree of intrasubjective stability of the vehicle required in order that these hoped for explanatory purposes be served is an issue for another day). So long as contentful states can be attributed to a range of subjects, we can hope for content-based psychological generalizations that apply to all of those subjects, regardless of what variations may appear, from subject to subject, in the intrasubjectively stable vehicles to which those contents are attached.

[31] One alternative is to identify the vehicles of content with vectors or regions in a state-space; see Evan Tiffany, "Semantics San Diego Style," *Journal of Philosophy* 96 (1999): 416-29; and Robert D. Rupert, "On the Relationship between Naturalistic Semantics and Individuation Criteria for Terms in a Language of Thought," *Synthese* 117 (1998): 95-131. Were the LOT theorist to state term-individuation criteria in the language of state-spaces, and vectors and attractors therein, an epigenetic argument parallel to the one given in this section would seem to lead to the same conclusion reached in text. Even in cases of many concepts normally classified as innate, *experience shapes* the relevant attractors; neither the attractors nor their relevant ancestors are present at birth, and thus, neither would seem to be the contents of those attractors, or of the components thereof (but see notes 37 and 38 below).

It is sometimes suggested that we individuate LOT terms by nonsemantically specified, functional criteria. In *Concepts*, Fodor would seem to have this approach in mind when he talks of modes of presentation (p. 19). However, individuation by functional role appears to lead to serious problems for Fodor, as a LOT-cum-causal theorist; see Murat Aydede, "Fodor on Concepts and Frege Puzzles," *Pacific Philosophical Quarterly* 79 (1998): 289-94.

[32] This assumption is, of course, controversial. For evidence of repeatable units that correspond to natural language terms and are susceptible to neurological characterization, see Friedemann Pulvermüller, "Words in the Brain's Language," *Behavioral and Brain Science* 22 (1999): 253-79; for suggestions as to how mental representations corresponding to linguistic elements develop epigenetically and as to how the nature of developmental interactions affects the character of the nonsemantically individuated mental representations of words, see Pulvermüller, op. cit., p. 263; Ralph-Axel Müller, "Innateness, Autonomy, Universality? Neurobiological Approaches to Language," *Behavioral and Brain Sciences* 19 (1996): 611-31, p. 629. For recent attempts to defuse what have become common multiple realizability-based objections to neural individuation of LOT terms, see John Bickle, *Psychoneural Reduction* (Cambridge: MIT Press, 1998), especially chapter 4; William Bechtel and Jennifer Mundale, "Multiple Realizability Revisited: Linking Cognitive and Neural States," *Philosophy of Science*, 66 (1999): 175-207; Robert W. Batterman, "Multiple Realizability and Universality," *British Journal for the Philosophy of Science* 51 (2000): 115-45; also see Mark Wilson, "What is This Thing Called "Pain"?--The Philosophy of Science behind the Contemporary Debate," *Pacific Philosophical Quarterly* 66 (1985): 227-67; more general concerns about multiple realizability can be found in the essays in Part II of Jaegwon Kim, *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993).

[33] See P.S. Churchland and T. J. Sejnowski, *The Computational Brain* (Cambridge: MIT Press, 1992) pp. 132, 307; D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, A. S. LaMantia, and J. O. McNamara (eds.), *Neuroscience*. (Sunderland, MA: Sinauer, 1997), Chapter 22.

[34] *Neural Darwinism: The Theory of Neuronal Group Selection* (New York: Basic Books, 1987); also see J. P. Changeaux, *Neuronal Man* (Princeton, NJ: Princeton University Press 1983/85, second edition, 1997).

[35] See, e.g., Steven R. Quartz and Terrence J. Sejnowski, "The Neural Basis of Cognitive Development: A Constructivist Manifesto," *Behavioral and Brain Sciences* 20 (1997): 537-96 (page numbers include peer commentary).

[36] Ibid., p. 581.

[37] One might take exception to this characterization of the state of developmental neurobiology, on the ground that selectionism seems to support a fairly strong nativism (see, e.g., M. Piatelli-Palmarini, "Evolution, Selection, and Cognition: From 'Learning' to Parameter Setting in Biology and in the Study of Language," *Cognition* 31 [1989]: 1-44): does not selectionism imply the presence at birth of a surfeit of nonsemantically individuated LOT terms, the bulk of which later disappear in a manner that accords with selectionist principles, the remaining terms having been "present from birth?" This would, however, misrepresent what is plausible in selectionist thinking; for first, a significant percentage of the human brain's synaptic connections develop after birth; this is a widely accepted view with which selectionists should take no particular exception--see Elman et al., op. cit., p. 291-94; for selectionist endorsement of this view, see Changeaux, op. cit. p. 248 (second edition); secondly, in most cases, experience-based reinforcement and pruning are required before whatever patterns of neural connections that exist at birth cohere as functional units. Thus, if we identify LOT terms, thought of nonsemantically, with functional cell assemblies, selectionism does not support the radically nativist view of nonsemantically individuated LOT terms: it does not imply that all, or even most, of the adult's nonsemantically individuated, primitive, unlearned LOT terms were present at birth.

[38] It is possible that content transfers from one nonsemantically characterized vehicle to another throughout the course of development (also, because plausible identity conditions for individual vehicles should allow changes over time in important physical properties of what continues to count as the same nonsemantically individuated unit, there may be some as yet unspecified sense in which the vehicle is present at birth--but see the preceding note). Perhaps the content of an innate concept, present at birth, attaches to one nonsemantically individuated LOT term, then "hops" to another as the next "host" develops. Recognizing this possibility, though, hardly makes it probable that there are present at birth vehicles carrying *all* the contents of terms typically claimed by cognitive scientists to be innate (or, more generally, of *all* of the primitive, unlearned LOT terms).

We should also consider the possibility that one could possess at birth knowledge regarding the application of a LOT term even when that term has not yet formed in one's cognitive or biological systems; however, without further development of this proposal, it is difficult to say whether it is at all plausible in the human case.

[39] Here Cummins may wish to follow some prominent nativists, dismissing the present discussion of neurobiological issues because of the paucity, to date, of substantive results in cognitive neuroscience. In Fodor's review of

Rethinking Innateness, he lambastes Elman et al. for assuming that certain neural mechanisms must encode representation: "[T]here isn't one, *not one*, instance where it's known what pattern of neural connectivity realizes a certain cognitive content, innate *or* learned, in either the infant's nervous system or the adult's " (*The London Review of Books*, Vol. 20, No. 12, 1997--reprinted as Chapter 12 of Fodor's *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind* [Cambridge: MIT Press, 1998], pp. 143-51; the quoted passage appears on p. 145). Chomsky expresses a similar view when resisting the demand that we cash out in neural terms nativist claims in linguistics; see "Language and Nature," *Mind* 104 (1995): 1-61. According to Chomsky, "The current situation is that we have good and improving theories of some aspects of language and mind, but only rudimentary ideas about the relation of any of this to the brain" (ibid., p. 11). Chomsky feels free, then, to make strong nativist claims in the absence of any supporting theory of neural implementation: "The conditions of language acquisition make it plain that the process must be largely inner-directed, as in other aspects of growth, which means that all languages must be close to identical, largely fixed by initial state" (ibid., p. 17). Compelling considerations, including poverty of stimulus arguments, imply innate content; it matters not whether we understand how precisely that arises from the subject's underlying physical being.

It seems to me that Fodor and Chomsky simply refuse to acknowledge recent decades' progress in cognitive neuroscience, although this is not the place to support my claim. To the extent that my claim is correct, however, Cummins should want to avoid endorsing Fodor's and Chomsky's positions *vis-à-vis* the relevance of neuroscience to our evaluation of concept nativism. Furthermore, the arguments of the preceding and following sections stand independent of any controversial claims about the relation between cognitive activity and development and their neurological bases.

[40] In presenting these arguments, it has been convenient to separate discussion of the emergence of content and questions about the emergence of the vehicles of content. However, the causal theorist might well suspect that these two processes are closely connected: in many cases, the very causal interactions with the environment that shape the brain's nonsemantically characterized representational resources at the same time contribute significantly to the fixation of the content of the units so shaped, so that, to a great extent, what does the shaping often exemplifies the kind to whose members the shaped entity comes to refer; see Rupert, op. cit., and Müller, op. cit., p. 629. Adopting this more integrated view of conceptual development would in no way weaken the arguments in the text, for the integrated view allows concepts to emerge in the dialectically pertinent ways: they can be unlearned, perhaps even

innate, emerging via postnatal interaction with the environment, all without the newborn's possessing (P2)-knowledge.

[41] How might such terms be coined? Changeaux suggests that protorepresentations (or something much like them--'prerepresentations,' he calls them) briefly cohere and pass away due to spontaneous, oscillatory firing of neurons (op. cit., pp. 139, 145, 169); experience with the world can then cause increased stabilization of certain prerepresentations, in keeping with the selectionist view summarized above.

Research that takes a dynamical systems-based approach to cognitive development suggests an alternative: Working within the framework provided by dynamical systems theory, Esther Thelen and Linda Smith propose that the child harnesses, by a process similar to trial and error, naturally generated "noise" to create new solutions, sometimes fairly stable ones, to the problems of cognitive and motor development. Although Thelen and Smith would almost certainly chafe at my referring to the harnessed units as 'terms in LOT', their view may in fact explain how the brain coins LOT terms, by transforming snippets of relatively chaotic processes into stable attractors in a dynamical system. See E. Thelen and L. B. Smith, *A Dynamic Systems Approach to the Development of Cognition and Action* (Cambridge: MIT Press, 1994), e.g., p. 134.

[42] There are two general ways in which we can conceive of the semantic status of such LOT terms as 'fur', 'legs', and 'whiskers', as these terms appear in the axioms associated with 'cat' (call these *auxiliary terms*): Either their content is in place at the time the axioms are framed, or they, like 'cat,' initially appear in the axioms as protorepresentations. When, in the text, I discuss explicitly represented theoretical axioms, I take all auxiliary terms to have their content fixed. A referee for this journal expressed concern that this assumption entails an unacceptable regress. Thus, in what follows, I explain in more detail the semantic status of auxiliary terms, and why their role in the coining of new LOT terms is unproblematic.

There are three routes by which auxiliary terms may have their content fixed prior to their appearing in the relevant axioms: either (1) by definitional combination of other LOT terms whose contents are fixed; (2) by the mediation of explicitly represented, but nondefinitional axioms in which auxiliary terms appear--axioms of just the sort presently under discussion, but directed at protorepresentations 'legs' or 'whiskers' instead of the protorepresentation 'cat'; or (3) by the mediation of implicit theory. Following Cummins (537), and in keeping with a broad consensus among philosophers of mind and language, I do not hope for much mileage from the first option: definitions are too few, the case of 'bachelor' notwithstanding; and in so far as auxiliary terms sometimes are fully

definable, we must explain how the contents of their definitional constituents were fixed. Eventually such definitional chains must be grounded in accordance with at least one of the other alternatives, (2) and (3). It is also possible that a given auxiliary term has its own set of explicitly represented, detection-mediating axioms, operative in the introduction of the auxiliary term itself. Of course, these axioms will contain their own auxiliary terms (e.g., auxiliary terms relative to 'whiskers'--such LOT terms as 'long' and 'thin'). Thus, as in the case of the first option, this story cannot be complete: it does not account for the ultimate introduction of LOT terms. For this reason, case (3) is of fundamental importance. Cummins allows implicit theory to play a detection-mediating role for at least some LOT terms, and the auxiliary terms 'fur', 'whiskers', and 'legs' may well be such terms (and if they are not, then other relevant terms will be: either the terms that make up their definitions, or the further auxiliary terms that appear in 'whiskers''s or 'legs''s explicitly represented, but nondefinitional, introducing axioms, and so on, as each case may be). It is not implausible to claim a significant stock of terms whose contents are fixed in accordance with possibility (3); to show this was part of the burden of the preceding sections. Surely this stock is not rich enough to serve as a definitional base for all other concepts, but that is not what we are after. We wish to find enough such terms to provide definitional constituents in cases where newly introduced LOT terms have definitions, but also, and this is surely the more common case, to provide the materials on which to draw when framing nondefinitional axioms that mediate the tokening of newly coined LOT terms. Recursive applications of such a process can yield a wide range of learned, primitive, nondefined LOT terms from a relatively small base of terms whose content is fixed via the mediation of implicit theory.

Rather than handling the threat of regress in the manner just described, the referee suggests treating auxiliary terms as protorepresentations *at the time they appear in the axioms in which such terms play their auxiliary role.* I do not rule out such a possibility. It may be that as a result of the kinds of process stressed above--maturation, dendritic arborization in response to environmental stimuli, etc.--an axiom consisting wholly in protorepresentations may take shape in the subject. However, if other protorepresentations wholly mediate the context-fixing detection-relations of a given protorepresentation *t*, the relevant axioms have the status only of an implicit theory of the things to which *t* will come to refer; for this is the dimension according to which we presently categorize a theory of *t* as explicit or implicit: whether the reliable tokening of *t* in response to the relevant distal properties or objects is mediated by explicit representations. In the present section, I set aside cases where mediating theory is implicit,

including cases where axioms consist wholly of protorepresentations, for here I intend to show how a LOT term can be acquired by the mediating effects of *explicit* theory.

It is possible that at least one of protorepresentation *t*'s axioms is partly explicit, partly implicit, containing at least one auxiliary term with content fixed and at least one protorepresentation beyond *t*. I trust that the views presented here provide a framework for handling such mixed cases, but a full discussion of them would lead us farther afield than constraints of space allow.

[43] Here one might imagine a best-match, or 'nearest neighbor', system with a threshold of difference beyond which no match will be made, and instead a new term coined; furthermore, the system could be such that when a match is made and the stimulus differs from the prototype, the stimulus "pulls" the prototype toward it in similarity space. For discussion of nearest neighbor, prototype-based categorization, see Churchland and Sejnowski, *The Computational Brain*, p. 76ff.

[44] That is, it may be a mistake to think that the cognitive system holds a theory as to when it will introduce a new LOT term (individuated nonsemantically) and when instead it will integrate the information given in new experiences into existing representational structures. In contrast, *I* surely am suggesting a theory (or, more accurately, a theory-outline) of the workings of one aspect of the cognitive system; see Cummins's discussion of this distinction in "Inexplicit Information," pp. 123-25.

[45] For statements of the orthodox view, see Fodor, *The Language of Thought*, pp. 31, 46, 56, 99; also see Fodor, *The Modularity of Mind* (Cambridge: MIT Press, 1981), p. 5; Fodor, *Concepts*, pp. 10-11; Pylyshyn, *Computation*, pp. 51-58; Churchland and Sejnowski, *The Computational Brain*, pp. 62-66. Whether the orthodox view grounds a criticism of the present proposal depends on how tightly one associates computation and semantically interpretable states. That is, one might claim that it is *essential* to computation that it only be carried out over semantically interpreted states or units; this understanding of computation seems to conflict with the present proposal. On the other hand, one might, as Pylyshyn does, think that our best cognitive theories will sometimes deal in computations over interpreted states, while at the same time including nonsemantic explanations of important aspects of cognition: "[W]e may need the biological level to explain other things as well, such as possibly the nature of cognitive development or maturation or psychopathology, and perhaps *some changes that are now called learning*; exactly what facts fall at each of the three levels [semantic, syntactic, and biological] remains to a large extent an open empirical question" "Computing in Cognitive Science," in M. I. Posner (ed.), *Foundations of Cognitive Science*

(Cambridge: MIT Press, 1989), pp. 51-91, p. 61, italics added; also see p. 72. I would extend Pylyshyn's remarks and claim that what sorts of state can be included in a computational model is also an empirical issue; if our most empirically successful computational models of cognition include dummy terms, with syntactic and biological properties, but no semantical ones, then so be it.

[46] On computational models of stereopsis, see David Marr, *Vision* (W. H. Freeman and Company: New York, 1982), pp. 111-59, and Churchland and Sejnowski, *The Computational Brain*, pp. 199-216.

[47] For a similar point, see Frances Egan, "In Defence of Narrow Mindedness," *Mind and Language* 14 (1999): 177-94, pp. 184-86, although note that Egan takes a more pessimistic overall view of the importance of content ascription than is assumed in the present paper. See also Chomsky's discussion of internalist computational views in "Language and Nature," especially pp. 51-56.

[48] Christopher Peacocke, "Computation as Involving Content: A Response to Egan," *Mind and Language* 14 (1999): 195-202, pp. 198-99; Fodor, *Representations*, pp. 22-23. Regarding the present proposal, it is unclear why any explanatory power is lost when we allow that some newly coined LOT terms participate in computations even though the terms lack content. This suggests a condition that must be met by a causal theory of content: it must assign content to newly coined, content-lacking LOT terms in a timely enough fashion that such terms possess content when our explanatory purposes demand that they do.

[49] It is sometimes claimed that semantics is prior to syntax for the following reason: What counts as a syntactic property in LOT is determined by the physical (or otherwise nonsemantic) properties causally operative in cases where we interpret the cognitive system in semantic terms (see, e.g., T. Crane, "The Language of Thought: No Syntax Without Semantics," *Mind and Language* 5 [1990]: 187-212). On this view, we must first characterize thought and behavior in semantic terms; then we can look for underlying formal or syntactic properties of, e.g., bits of the brain, that will explain the occurrences of semantically interpreted behavior. Here I take no issue with this view, for it is consistent with the present proposal: once we have discovered the operative formal properties--those that explain cognitively characterized skills and capacities--we are free to identify further instantiations of these formal properties as LOT terms, whether or not these terms yet play semantically characterized roles in the cognitive system.

[50] One might wonder, as did a referee for this journal, whether the present defense of CT is consistent with the nativism of one of CT's most ardent advocates: Jerry Fodor. Throughout the paper, I have emphasized ways in

which the subject's interaction with her environment shapes her cognitive resources, and thus the views presented here may seem squarely nonnativist. The apparently nonnativist implications of the present proposal may be appearance only, however; in so far as Fodor counts a concept as innate if and only if that concept is not acquired by a rational-causal process, the present view might in fact be consistent with Fodor's extreme nativism; much depends on how one conceives of the axioms discussed above, whether, for example, they consist in hypotheses about the meaning of a newly coined LOT term, or whether their constituent auxiliary terms play only a brute-causal, triggering role. No doubt we should like to know (a) whether the present view implies the innateness of any LOT terms--by a Fodorean criterion or any other prominent one--and (b) if so, which LOT terms would thereby be classified as innate. Nevertheless, given the complexity of such issues, limitations of space prevent their satisfactory prosecution here.

[51] A referee for this journal raised the concern, as did Edward Averill, that my defense of CT coheres poorly with an atomistic version of CT, such as Fodor's. Because a newly coined protorepresentation appears in axioms alongside full-fledged representations, one might think that, even though the axioms do not provide definitions, the new LOT term inherits its content from its relations to these full-fledged representations. Bear in mind two points: (a) Fodor allows intentional mechanisms to mediate the causal relations that atomistically fix content for LOT terms; atomism is preserved because the ultimate ground of content remains the causal relations themselves, nonsemantically characterized (*Psychosemantics*, p. 121, and *Concepts*, pp. 78-79); in fact, Fodor allows that in the human case, there may be certain concepts we cannot acquire without mediation by intentional mechanisms, but Fodor does so without abandoning his semantic atomism (*Concepts*, p. 157-58) (for further discussion of these points, see my "Dispositions Indisposed: Semantic Atomism and Fodor's Theory of Content," *Pacific Philosophical Quarterly* 81 [2000]: 325-48) and (b) the present defense of CT is intended only partly as a defense of Fodor's views; my primary intent is to defend the version of CT I have proposed elsewhere (see Robert D. Rupert, op. cit.; "The Best Test Theory of Extension: First Principle(s)," *Mind and Language* 14 [1999]: 321-55; and "The Best Test Theory of Extension" [Ph. D. Dissertation, University of Illinois at Chicago, 1996]) (this point should also be kept in mind when considering the issue raised in footnote 50 above). In developing my own views about content, I have been heavily influenced by Fodor's work; nonetheless, the version of CT I offer differs significantly from Fodor's. Regarding the present concern, that mediation by intentional mechanisms fixes content nonatomistically, I would, following Fodor, emphasize (a), at least with respect to many learned, primitive LOT terms. In certain cases,

however, I incline toward allowing a greater degree of holistic (or molecular) content-fixation than does Fodor; why I do so is an issue to be taken up elsewhere.