CHAPTER 14

INNATENESS

STEVEN GROSS AND
GEORGES REY

To what extent are the structures and contents of the mind innate, and to what extent are they learned or otherwise acquired from the environment? Versions of this question have shaped theorizing about the mind since the ancient Greeks and continue to divide researchers today. The debates concern a wide range of traits—for example, the capacity for color perception and discrimination, the ability to follow a gesture, and even a penchant for surprising people by sneezing in elevators (Bouchard et al. 1990). Some touch upon matters of great general concern with possible implications for public policy, such as the nature of IQ, gender preference, and criminality. The relevant empirical methodologies are increasingly complex and varied: meta-statistical demographics, twin studies, examination of phenotypic correlations, study of early childhood development, etc. The debates concern not just what is innate, but also the prior questions of what innateness is, whether there is just one scientifically legitimate conception of innateness, and indeed whether this is any at all. Accordingly, we examine below several attempts to articulate a conception of innateness that could do some explanatory work (Section 2). We then focus on the philosophically salient case of whether *ideas*, or *concepts*, are innate (Section 3).

The topic of innateness has a rich history, an understanding of which illuminates the contemporary debates on which we will focus. We thus preface our discussion with some brief historical remarks (Section 1).

1. Historical Background: Rationalism and Empiricism

The earliest and most famous argument for conceptual innateness occurs in Plato's *Meno*, where Socrates purports to demonstrate that a slave boy, who has received no explicit instruction, has sufficient ideas somehow available to him to understand a nonobvious proof in Euclidean geometry. Socrates concludes from this demonstration that the slave boy must be "recollecting" the relevant ideas from a previous incarnation. Plato's student, Aristotle, reacted against this suggestion with what is perhaps the first empiricist proposal. On his view, *all* ideas are derived from experience by a causal process in which "forms" (or properties of things) in the external world are transmitted into the mind (Aristotle 1968, 417–26). With the medieval resurgence of interest in Aristotle, the view was defended in highly influential writings of Aquinas (1266/1948, I, 87, a, 3; see Adams 1975, 73–74 for discussion), and is arguably a chief component of much of what passes today as common sense on the topic.

In the modern period, John Locke (1690/1975) also defended a strong form of empiricism about concepts, insisting that our simple ideas are derived from sensation, and all other ones are constructed from the simple ones by the mental operations of "compounding," "comparing," and "abstracting" (Book II, ch. 12). Interestingly, he also maintained that our ideas of at least secondary qualities, such as color and sound, are *not* caused by those very properties in the external world (Book II, ch. 8), inviting the suggestion, widely presupposed ever since, that at least these elementary perceptual ideas are innate in us (see Fodor 1981, 275–77 for discussion).

Full blown nativist proposals resurfaced in the modern period in the work of Herbert of Cherbury (1624/1937), Ralph Cudworth (1678/1999), and, most famously, René Descartes (1647/1911). Descartes is especially impressed by the fact that the geometric figures studied by mathematics are not physically possible objects of sensory perception (1641b/1970, 227). Even so basic an idea as that of an *enduring substance*, such as a portion of wax that seems to us to remain constant as it undergoes various physical transformations, seems to require an idea of *substance* that experience alone cannot provide (Descartes 1641a/1970, Meditation 2).

Descartes emphasized that the issue concerns not *occurrent*, but *dispositional* properties of a neonate—a point relevant to our discussion below. No one thinks that infants are born actually *entertaining* the full panoply of geometric ideas that might be innately available to them. Rather, innate ideas are like innate character traits and diseases, which people "are born with a certain disposition or propensity for contracting" (1647/199, 442), and which may require appropriate circumstances to be activated (442–43).

Following upon Darwin, nativism about "instincts" and other cognitive mechanisms received some support in psychology (James 1890/1983, ch. 24) and ethology (Lorenz 1957; and, for critical discussion, Lehrman 1953, 1970). But empiricism became the orthodoxy in at least Anglophone philosophy and psychology until the revival of rationalist approaches to the study of language in the work of Noam Chomsky (1965, 1966, 1968/2006) (see Chapter 15). Advancing what has come to be called a "poverty of stimulus" argument, he called attention to the inadequacy of the data to which children are standardly exposed to determine the elaborate grammars that they quickly, effortlessly, and universally acquire, an observation that cognitive scientists have since applied to many other domains, such as the understanding of objects, number, animals, artifacts, minds, and morals. (A range of contemporary nativist arguments and claims concerning cognition can be found in Mehler and Dupoux (1994) and Carruthers, Laurence, and Stich (2005, 2006, 2007).) The revival of *conceptual* nativism more specifically is most associated with the work of Jerry Fodor, which we will discuss at length in Section 3.[1]

2. What is Innateness?

But just what is at issue in innateness debates? Might it be, as some claim (e.g., Griffiths 1997), that we lack any scientifically legitimate conception of innateness at all? If so, are innateness debates in fact empty?

We can only focus on a few representative views here (see Mameli and Bateson 2006, for discussion of twenty-seven candidate conceptions of innateness). We begin (Section 2.1) with proposals that would subsume the innateness of psychological traits under a more general positive account of innateness in biological terms. We then turn (Section 2.2) to a proposal that focuses on *psychological* innateness, characterizing it negatively in terms of how innate psychological traits are *not* acquired—in particular, not by learning. Finally, we argue (Section 2.3) that, even if these proposals fail to sufficiently articulate scientifically legitimate conceptions of innateness, this need not undermine innateness debates—in particular, those concerning concepts, the topic of Section 3.[2]

[1] Most contemporary treatments diverge from many of the historical ones in not linking the issue of innate ideas to any claims of *a priori* knowledge, or knowledge justifiable independently of experience. It is now widely presumed there could be plenty of innate beliefs that are false (e.g., that space is Euclidean). Cf. Hart (1975).

[2] Our concern is not with the meaning of the vernacular term *innate*, or with folk conceptions of innateness. The question is what innateness may be, and what may be innate, from the perspective of our best current scientific theorizing. We nonetheless mark proposals' counterintuitive consequences. For, with sufficiently significant departures (for example, if what is learned were allowed to be innate), one may wonder whether we retain a conception of *innateness* at all—or at least whether it is advisable to retain the *word* "innate."

Griffiths (2002) suggests a distinct reason for avoiding the word. He maintains that the vernacular conception of innateness is an expression of a largely unreflective and automatic

2.1. Biological Conceptions

In the light of modern biology, it is natural to interpret nativism in terms of the contribution genes make to a trait's emergence. But a satisfactory formulation of this idea has proven elusive. Approaches that emphasize causal determination face the problem that prima facie practically all traits (among them innate traits) result from the *interaction* of genes and environment (that is, anything not part of the genome). Even the emergence of so seemingly paradigmatic an innate trait as eye color depends upon intricate genetic-environmental interactions. Moreover, it is unclear how to apportion causal responsibility. As Sober (1988) emphasizes, there is no "common currency" (312) with which to compare the relative contributions of genes and environment; unlike physical forces, biological determinants do not in general decompose into amounts of genetic versus nongenetic "force." Some other approach would be needed to identify when the genetic contribution is appropriately "critical" in a way that can underwrite a claim of innateness (cf. Waters 2007).

It might be thought that technical notions of *heritability* employed in population genetics might be of use here, where the genetic heritability of a trait within a population is *the proportion of phenotypic variation due to genetic variation*. But the heritability of a trait is not even defined for populations in which there is *no* variation—as with, for many populations, the intuitively innate trait of *having a head*. Moreover, intuitively innate traits—such as having five fingers—can exhibit low heritability: the majority of people lacking five fingers may be victims of accidents.[3] Another strategy would borrow ideas employed more generally in attempts to naturalize content (see Section 3.4.1 below) to cash out genetic determination in terms of a trait's being "coded" for in the genes. But scaling up such suggestions beyond the representation of amino acid triplets and proteins remains problematic (Godfrey-Smith 2007).

Alternatively, one might avoid the problem by rejecting the claim that *innate* traits result from the interaction of genes and environment—more specifically, by identifying innate traits with the properties of the genotype itself. If one does not require that such traits be characterized in molecular or non-relational terms, this identification can be less restrictive than it might appear at first, for it then can include dispositional properties, such as the disposition to have concept C activated

folk-biological essentialism that "acts as a sink…draw[ing] new, stipulative usages back towards the established use" which, confounding empirically dissociable properties, encourages illicit inferences. (Folk biology is itself significantly innate according to some, though not Griffiths. Cf. Medin and Atran 1999.) For some initial empirical investigation into the vernacular conception of innateness, see Griffiths et al.(2009). Note that, where they and others speak of concepts, we speak of *conceptions* so as to avoid prejudging whether elements of a conception are constituents of a concept—cf. Samuels (2007). For a pragmatic argument against eliminating either the term *innate* or the concept of innateness from scientific discussion, see Cowie (2009).

[3] The conflation of genetic heritability, genetic determination, and innateness has muddied public debate concerning IQ in particular. Cf. Lewontin (1974), Block (1995), and Sober (2001). For a discussion of *non*genetic heritability, see Mameli (2004).

in appropriate circumstances (see Section 3). The restriction, however, precludes the possibility of innate traits that emerge in development: for example, having a head or secondary sexual characteristics (including their psychological aspects), as opposed to the disposition to develop such traits.

Some of these suggestions are worthy of further exploration. But for the remainder of this section we concentrate on two versions of another approach. These grant the indispensability of environmental contributions to the acquisition of innate traits, but emphasize the irrelevance of environmental variation.

2.1.1. *Invariance*

Sober formulates his "Invariance" conception of innateness as follows:

> (INV) a phenotypic trait is innate for a given genotype if and only if that phenotype will emerge in all of a range of developmental environments.
> (1988, 795).[4]

Such traits are said to have a "flat norm of reaction" for the genotype: plotting the trait as a function of the relevant environments yields a flat line. Invariance is easily confused with species universality. But where the former asks what *one* genotype would yield in varying environments, the latter asks what *various related* genotypes have in fact yielded in actual environments.

(INV) faces two major challenges: (i) to specify the relevant range of environments (a recurring problem, as we will see), and (ii) to accommodate invariance apparently owed to environmental stability. A third, more theory-laden challenge would be to accommodate innate traits that are not invariant because they are only present when triggered (see Section 2.2.2).

The basis of the first challenge is obvious enough: absent some restriction on the relevant range of environments—for example, to exclude conditions of extreme deprivation or radical intervention—practically no trait is innate according to (INV). But it is not an easy task to provide a specification with an independent scientific basis.

If we restrict the range to environments that are or have been statistically typical, we face the problem that many genotypes are tokened but once (note that a whole individual's genome tokens a specific genotype): their actual environment is thus their typical environment, rendering all their traits innate. If we advert to what is or has been typical for their species, there is the problem that environments intuitively extreme for one morph might be typical for another.[5] Appealing to the

[4] Cf., for example, Tooby, Cosmides, and Barrett (2005, 323, fn. 7): "What we mean [by an innate trait] is that it reliably develops across the species' normal range of environments." Also, Goldin-Meadow (2003, 215): an innate trait is something "whose development is, if not inevitable, certainly one that each organism in the species is predisposed to develop in widely varying circumstances"—though note our remark below on species universality.

[5] The formulation "what is or has been" includes proposals that advert to what evolutionary psychologists call the Environment of Evolutionary Adaptiveness. Cf. Tooby and Cosmides (1990).

typical environment of some smaller set of genotypes would not address the further problem that some intuitively innate traits in fact emerge in very few environments: most spiders fail to reach adulthood. Conversely, the occurrence of certain statistically rare environments seems to render some traits *non*-innate: bees differentiate into workers and queens depending on what they are fed in the larval stage, but very few bees are exposed to a queen's ration. We might instead try restricting the range to what is "normal" in some nonstatistical sense, but this only heightens the demand for an independent scientific basis for this restriction—an issue to which we return in Section 2.2.

Sober (1998) suggests that there might not be a single specification of relevant environments: one might need to fix the range "pragmatically" (795), presumably as it varies with (scientifically legitimate) explanatory interests—different interests picking out different ranges of environments, much as different interests might pick different ranges of circumstances to identify something as a solvent or a poison. If different interests could be in play on different occasions regarding one and the same trait, then innateness would be a relative property—or, alternatively, the term *innate* would express different properties in different contexts.

The second challenge is that, on any reasonable restriction of the environmental range, there seem to be intuitively non-innate invariant traits. Consider a person's belief that she has a nose. This belief is presumably learned and so not innate. But arguably this belief would have emerged in any of the environments relevant to (INV) (cf. Stich 1975b, 9; and Wendler 1996, 92–94). Sober's pragmatic reply to the first challenge seems only to heighten the difficulties here, since practically any trait will be innate at least relative to some range if there is some explanatory interest that has us hold it fixed (suppose we want to know why not everyone who learns grade school math can master the calculus).

A standard diagnosis of this difficulty is that (INV) fails to place proper constraints on the process by which a trait is acquired. It asks only whether it always *would* be acquired without distinguishing the roles of endogenous and exogenous contributions. But it is unclear how best to remedy this lack. Mallon and Weinberg (2006) add a requirement that the process be "closed" in the sense of normally leading to one outcome. But some learning-like processes are closed in this sense—for example, some forms of imprinting, as when some species of parasite normally imprints on a particular type of host (Mameli 2004 discusses some cases, albeit for a different purpose). An alternative attempt, to which we now turn, is Ariew's (1996, 1999, 2007) "Canalization" proposal.

2.1.2. Canalization

The term *canalization* was coined by the biologist C.H. Waddington (1957, 1975) to refer to a trait's relative insensitivity to genetic and environmental perturbations. The label comes from his comparison of an organism's development to a ball rolling down a grooved landscape. The grooves (or canals) represent the organism's genetically determined developmental possibilities. A trait is genetically well-canalized to the extent that genetic variation would not affect the canals that channel development in its direction. We can extend the metaphor to capture *environmental* canalization by allowing the landscape's topography to be determined in part by environmental factors as well. A trait then is environmentally well-canalized to the extent that changes to the *environment* would not affect the canals that channel development in its direction. Ariew proposes that we deploy Waddington's notion of environmental canalization to characterize innateness:

> (CAN) A trait is innate for a genotype to the degree to which its development is "insensitive to a range of environmental conditions." (Ariew 1999, 128)

The crucial difference between this proposal and (INV) is the appeal to "insensitivity." But how is this to be understood? It would not suffice to let talk of insensitivity simply mark the relevant difference between endogenous and exogenous invariance: the question in the first place was whether there is a scientifically legitimate way of articulating this distinction. Moreover, Ariew (1999, 123–26) follows Sober in emphasizing that there is no factoring out the comparative causal contributions of genes and the environment. At least in this sense, practically all traits are sensitive to both sorts of factors without one kind of factor intelligibly playing a larger or more significant role than the other.[6] If there is some other relevant sense of sensitivity, it must be supplied.

Some alternative attempts to cash out "insensitivity" come at a cost. Suppose, for example, that a trait's development is insensitive to the environment if it not only invariantly emerges, but moreover emerges in an invariant way—suggested perhaps by Collins's (2005, 167) discussion of "developmental implasticity." So construed, (CAN) clearly differs from (INV): if various developmental pathways can lead to one's believing one has a nose, then according to (CAN) it is not innate. But it is unclear why intuitively innate traits must have invariant developmental pathways (however such pathways are individuated). Many genotypes exhibit genetic redundancy: some other gene or genetic pathway can compensate for the inactivation of a gene otherwise central to the development of a trait. Waddington's epigenetic landscape in such cases would contain multiple branching canals that at some point flow back together toward their shared phenotypic goal. Why should the existence of "backup" developmental pathways preclude innateness? Similarly, if the backup involves learning: why should the *possibility* of acquiring a trait through learning undermine the innateness of a trait that as it happens is not learned? Consider the species of canary Ariew (2007, 572) mentions elsewhere that can acquire its song either through learning or hormonal triggering.

These problems suggest another reading of (CAN), albeit one disavowed in Ariew (2007). Perhaps so far as innateness is concerned, what is crucial to how a trait develops is not that it develop in just one way, but that the developmental

[6] Mechanisms that arguably have evolved specifically to "buffer" a developmental pathway against environmental perturbations (see below) can work precisely because they *are* sensitive to the environment in this sense.

pathway that is in fact followed be *of the right kind*. In particular, it must involve a mechanism that evolved (or was co-opted) to have the function of "buffering" development against certain environmental contingencies. But here too we face several problems. First, we are forced to count genetic disorders as non-innate. Second, we are forced to count traits as non-innate if their presence is explained, not by the functions of evolved mechanisms, but by developmental or lower-level physical constraints, as in Cherniak (2005) on optimal neural wiring, and Chomsky (2005, 2007) on computational efficiency.[7] Finally, it is unclear that (CAN) so construed rules out learned traits: a trait could be canalized to be learned or otherwise acquired from experience. As Mameli and Bateson (2006, 172) point out, Sterelny (2003) argues that our folk biological ability to taxonomize animals is canalized in the current sense but is nevertheless in part culturally acquired.

This last worry recalls the objection to Invariance based on intuitively non-innate but invariantly acquired traits—for example, invariantly learned beliefs. Ariew's (2007) reply to such cases involves adducing environments in which the relevant environmental factors are absent. But these environments are arguably abnormal, and intuitively innate traits can be rendered non-innate in the same way: since both innate and invariant but non-innate traits have developmental pathways that would be disrupted in certain environments, a basis is needed for treating cases differently.

Ariew's response (2007—but see also 1999) is that what matters for innateness is whether a trait's emergence is sensitive to *certain specific kinds* of environmental factors, where the relevant factors can vary with the trait in question and indeed with one's explanatory interest. It is this that should fix the relevant range of environmental variation. Thus, *some* birdsong is innate because its emergence does not depend on song-like acoustic cues, and the Language Acquisition Device is innate insofar as its emergence does not depend on linguistic input—even if, as Ariew emphasizes, both depend on environmental contributions for emergence at all.[8] The belief that one has a nose, however, *is* dependent on particular sorts of experience relevant for assessing its innateness; so it is legitimate to consider an environmental range lacking such experiences—even if in some cases the environment is only "conceptually possible" (2007, 577). What gives content to Ariew's account then is not the particular contribution genes make—despite occasional remarks such as that a canalized trait is one whose development is under "strict genetic control" (1999, 134). Rather, it is a sense of what kinds of environmental contributions to disallow in particular cases—one presumably based on the explanatory success of particular developmental models and strategies. (See Griffiths 2009, section 5,

[7] It is not even clear whether *genetic* canalization is the result of selection pressures as opposed to, for example, genetic-developmental constraints. See Siegal and Bergman (2002) and Wilkin (2003).

[8] Ariew simplifies his examples in order to illustrate the idea. For example, the innate birdsong to which he refers remains highly schematic, although recognizably similar, in comparison to the normal song of conspecifics who have received the relevant cues; see Gould and Marler (1991).

however, for skepticism that research in developmental biology, including on birdsong, in fact vindicates the usefulness of such a conception of innateness.)

Arguably, Ariew's view mainly differs from Sober's in that it says more about the basis for pragmatically varying the relevant environmental range. Cashed out in terms of the exclusion of *specific* environmental factors, canalization also bears a close relation to the next conception we examine.

2.2. Psychological Primitivism

Fiona Cowie (1999) and Richard Samuels (1998, 2002, 2004) develop a conception of innateness that, like Ariew's, is designed to improve upon (INV) by factoring in the way an organism comes to possess a trait. There is a surface difference in that Ariew presents his characterization of innate traits positively in terms of how they *are* acquired (via a canalized developmental pathway), while Cowie and Samuels's Primitivism characterizes them negatively in terms of how they are *not* acquired. This difference vanishes, however, if canalization must be cashed out in turn by reference to excluded environmental interactions. Primitivism then differs first and foremost in that it is restricted to a particular domain and that it locates its exclusion of particular environmental interactions at a disciplinary boundary. Thus Segal (2007) suggests that the two strategies simply reflect two sides of the same phenomenon, at least so far as psychological traits are concerned (more on this below).

Specifically, Primitivism builds on the thought that innate traits are not *learned*. Of course, not being learned, even if necessary for innateness, intuitively does not suffice—consider sunburns. But the suggestion is that, suitably refined, it might suffice for the innateness of *psychological* traits. The idea is that, however such traits are acquired, so long as they are not acquired via a psychological process, psychologists can treat them as primitives so far as their own theorizing goes—and this is what is at issue at least in (many) innateness debates in the cognitive sciences.

Cowie's concern—so far as primitivism as a conception of innateness goes—is to distinguish it from other conceptions and to identify a tradition based upon it that connects early modern debates with Fodor's work. Her critical discussion concentrates more on Fodor's claims concerning *what* is innate, so understood. Samuels's discussion, on the other hand, spends more time worrying whether this conception is itself scientifically viable; for that reason, we focus here on his development of the position (aspects of Cowie's critical discussion come up in Section 3 below).

2.2.1. Primitivism and Overgeneration

Samuels's initial statement of primitivism is as follows:

> (PRIM)…a psychological structure is innate [for a genotype] just in case it is a psychological primitive.…[i.e.,] a structure posited by some correct scientific psychological theory [but such that] no correct scientific psychological theory…explains [its] acquisition…(2002, 246)

What counts as a correct scientific psychological explanation of acquisition—Samuels's potentially less restrictive substitute for "learning"—is for science to say. But Samuels mentions explanations that advert to perception, inference, or conditioning. Nonpsychological explanations would include neurobiological and molecular biological explanations that do not advert to such processes. It would include as well "brute-causal" triggering by external factors, including triggering that follows upon experiential input, a case particularly important for discussions of concept acquisition (see Section 3.1.2). The distinction is no doubt unclear: for example, it is unclear what subpersonal computations count as inferential. But perhaps it is a virtue of the view that it reflects the crux of many first-order debates about psychological innateness, and thus locates what future research must clarify if this conception of innateness is to prove legitimate.[9]

Samuels worries, however, that his account overgenerates, admitting intuitively non-innate psychological traits whose acquisition is not explained psychologically. Head trauma, stroke, or surgery, for example, can cause brain lesions with a variety of psychological consequences (e.g., altered personality, memory, problem-solving capacities, etc.). In one well-known case, the mind of Phineas Gage, a nineteenth century railway worker, was dramatically transformed by an iron rod that passed through his frontal lobes after he accidentally tamped it into fused gunpowder (Damasio 1994). With good reason, Gage's friends reported that he was no longer the man he was, not that his innate self had been triggered.

A possible reply is that closer examination would reveal a psychological component to such cases. The thought is not that a psychological process might intervene between neural cell death and psychological end state (though that might indeed cover some cases). Rather, it might be that Gage, for example, wound up as he did in part because of earlier psychological acquisition processes: arguably, his psychological end state resulted from that sort of cell death occurring in *that* sort of mind/brain.

This reply has the prima facie virtue of potentially helping as well with at least some of a wide range of more mundane cases that Samuels does not discuss. Perceptual states seem to brute-causally yield non-innate psychological states in a variety of ways: consider the emotional effects of music or a warm bath, or the experience of "love at first sight." Learned associations might provide a reply in some cases. A broader list of psychological acquisition processes might cover others—so long, of course, as the processes were reasonably distinguished from brute-causal triggering. But, for some, one might need to invoke earlier psychological acquisition processes that prepared the way for the mundane cases at issue.

On the other hand, the reply raises delicate questions of individuation: just what should be included in a trait's acquisition process? Gage would not have

[9] Cf., for example, Carey (2009, 453):

> "Innate" simply means unlearned—not the output of an associative process, a hypothesis testing mechanism, or a bootstrapping process—that is, not the output of any process that treats information derived from the world as evidence.

acquired those traits in the way he did had he not learned how to set a fuse in gunpowder. But presumably his having learned *that* does not suffice to render the acquisition process psychological. Note also that our more mundane cases all involve perception, explicitly deemed a psychological acquisition process by Samuels. So, one is tempted to count the emotion's acquisition process psychological as well by including the perceptual process as a part. But what would warrant doing so in such cases, but not in cases involving perceptual triggering? (See also the discussion in Section 3.4.2 below concerning the role learning a stereotype may play in concept acquisition.)

2.2.2. *Normal Development*

In any event, Samuels pursues a different reply to the overgeneration problem. He maintains that the cases that worry him—rods through heads, strokes, and presumably at least some forms of medical intervention—are clear instances of abnormal development, so he adds to (PRIM) a normalcy clause that would exclude them:

> (PRIM*) A psychological structure is innate for a genotype just in case it is a psychological primitive and would be "acquire[d] . . . in the normal course of events." (2002, 259)

Of course, the added clause does not address our mundane, clearly not abnormal cases—nor is it intended to—so they would have to be handled some other way, perhaps as above.

(PRIM*) might be read as requiring either (i) that the trait not be acquired abnormally (i.e., that it would emerge in *a* normal course of events) or (ii) that it would emerge in *all* normal courses of events (so that Samuels's view becomes a combination of unrefined Primitivism and Invariance restricted to normal environments). Questions concerning triggering arise either way.

The first reading has problems with some nonpsychological modes of acquisition capable of yielding different psychological traits in different normal environments. Clear cases are difficult to supply in the absence of empirical details, but the conceptual point is clear enough. Suppose that, along lines of Chomskyan linguistics, the specific grammatical rules a speaker respects are determined by the setting of certain "parameters," such as whether verbs precede or follow their objects (an "SVO" versus an "SOV" language), and these parameters are triggered by certain stimuli. Then one's speaking an SVO language as opposed to an SOV language would, implausibly, count as innate.[10]

The second reading would preclude innate psychological traits that happen to be triggered in only *some* normal courses of events, as one might hypothesize of,

[10] The supposition that grammatical parameter setting is a matter of triggering, however, is contrary to fact if it involves discerning statistical patterns, and if discerning statistical patterns, at least in the way the developing language faculty does, counts as psychological (cf. Yang 2004; Scholz and Pullum 2006).

say, some mathematical or musical ability. Of course, Samuels may diverge from intuition, but on this point, for better or for worse, his view would diverge as well from his understanding of Fodor's conceptual nativism—and Samuels holds that, all else being equal, an account of innateness should "preserve the standard categorization of central figures" (2002, 239). On Samuels's understanding of Fodor, concepts are innate because they are acquired by triggering. But such triggering need not invariantly occur: indeed, Samuels raises just this as a further objection to Sober's Invariance proposal. This divergence from Fodor is avoided, however, if the triggering is rather understood as only activating a concept already possessed—cf. Section 3.[11]

In addition, further refinement would be needed to accommodate traits that can be acquired in multiple ways, as discussed in Section 2.1.2. (PRIM*), for example, would not exclude abnormally acquired psychologically primitive traits that *would have been* acquired psychologically in some or all normal environment(s). This complication can be avoided by dropping Samuels's modal language: we might simply require that the primitive not *in fact* be acquired abnormally. But if there are triggered, non-innate psychological traits (the SOV example), then again we would need to consider as well what would emerge in other normal environments, contrary—as we just saw—to what can be allowed by conceptions that admit non-invariantly acquired innate traits.[12]

On either reading, if (PRIM*) is to articulate a scientifically legitimate conception of innateness, it is important that its notion of normalcy be neither evaluative—based on some conception of how we ought, or are supposed, to be—nor merely reflect a "folk" conception of our proper environments or course of development.[13] But there is room to question whether a scientifically legitimate alternative is available. We have already mentioned reasons for not adverting to what is statistically typical. Nor does a functional conception, invoking environments in which an

[11] Fodor (1998) later retreats from a commitment to concept nativism—see Section 3.4.2—allowing that concepts can fail to be learned but not thereby be innate, even in normal cases. Samuels's divergence from Fodor on *this* score might be mitigated by allowing for a relativization of innateness to specific developmental systems (discussed below): the concepts could be primitive for psychology, but not relative to some other developmental system of which they are the outcome.

[12] If there are innate genetic psychological disorders (perhaps autism), then their actual course of development must count as normal in the sense used here. This might sound counterintuitive, but perhaps only if one conflates a notion of genetic abnormality with what environments are normal for the genotype. That said, it may be unclear what should count as normal for an abnormal genotype.

[13] The evaluative sense is of course important and (we hope) guides our attempts to *improve* our environments—medically and otherwise. But this does not suffice for it to scientifically legitimate Samuels's conception of innateness. Consider Phenylketonuria (PKU), a genetic disorder that leads to mental retardation unless both the mother when pregnant and the child afterwards adhere to a severely restricted diet (cf. Kitcher 1996). If only "restricted diet" environments count as normal, then, according to (PRIM*), PKU-mental retardation is not innate and its absence is; if both

organism lives and thrives, seem to help. The natural nonevaluative explication would be in terms of fitness—so that normal environments were ones, roughly, that led to more offspring. But medical intervention (e.g., personality-altering brain surgery) can increase fitness.

Samuels does not expand on the notion of normalcy. In defense, he notes that *ceteris paribus* clauses that assume "some largely unarticulated set of normal conditions" (2004, 140) are a common feature of all sciences with the possible exception of physics, so that an appeal to normalcy in this case raises no special problems. Indeed, such an appeal may not be to any specific set of normal conditions so much as an exclusion of apparent exceptions to a law as due to independent interference (cf. Pietroski and Rey 1995; and, for general discussion of *ceteris paribus* laws, Earman, Glymour, and Mitchell 2003). But even if such a notion raises no special philosophical problems, it does not follow that it is empirically legitimate in this case. Like his psychological/non-psychological distinction, Samuels's normalcy clause places an empirical bet: that science will indeed find it fruitful to consider all of those developmental factors abnormal interferences. It is clear enough that we often have an interest in how the occurrence or absence of an accident, surgery, or stroke would affect outcomes. But then we are also interested in how the occurrence or absence of various dietary, genetic, educational factors, etc. would affect outcomes. What is not clear is that there is an independent *scientific* reason for deeming the former abnormal. It is perhaps noteworthy that, again, developmental biologists are arguably among the most skeptical that there exists a scientifically useful notion of innateness (cf. Griffiths and Gray 1994; Oyama 2000; Johnson 1997; Bateson 2000).

2.2.3. Generalizing Primitiveness

Psychological Primitivism expressly limits its ambitions. Some might consider this a liability, others, a positive asset. In any event, it is worth noting, first, that nothing in Primitivism precludes the possibility of other scientifically legitimate conceptions of innateness. It is, for example, in principle consistent with any of a variety of biological conceptions (we return to Canalization in particular presently). But

restricted and unrestricted diets count, then neither presence nor absence is innate on reading one above, but both are on reading two; if only unrestricted diets count (what until recently would have been typical, perhaps universal), then PKU-mental retardation is innate but not its absence. It is clear what environments we want, but not clear what environments are abnormal in a (nonstatistical) scientifically useful sense. Similarly for other environmental alterations and innovations it is "normal" for us cultural creatures to introduce: clothing, improved diets, correction of vision, dental care, types of ornamentation, etc.—many of which have psychological consequences.

Regarding folk conceptions, note that Griffiths (2002) and Griffiths et al.(2009) maintain that the vernacular conception of innateness includes a nonscientific notion of *intended outcome*: "how the organism is *meant* to develop [so that] to lack the innate trait is to be malformed [and] environments that disrupt this trait are themselves abnormal." (2009, 609)

Primitivism does not require this: it could capture what is at stake in innateness debates concerning *psychological* traits even if biological innateness proved a chimera. It is also possible that biological innateness (assuming it is not a chimera) and psychological innateness as the primitivist conceives it could diverge—for instance, if an unlearned psychological trait were deemed biologically non-innate owing to the particular role of environmental factors such as diet (such considerations get raised in debates concerning, e.g., IQ and autism). Of course, the possibility of multiple legitimate innateness conceptions poses no problem apart from the risk of terminological confusion (cf. Cracraft 2000 on different conceptions of "species").

Second, the primitivist strategy might generalize, so one could view *Psychological* Primitivism as an instance of a more general template for generating distinct innateness conceptions. Psychological Primitivism isolates *psychological* traits and identifies those that are innate based on the absence in their normal acquisition of certain processes proprietary to the psychological: processes that involve proprietary interactions with, or input from, the environment. One might *mutatis mutandis* likewise identify innate traits in other domains. In famous work for which he received a Nobel Prize, Niels Jerne (1985) postulated that all the antibodies people ever develop in reaction to disease are already available prior to exposure, waiting to be activated by a pathogen. Whatever else might be true of their acquisition, such antibodies can be said to be innate in the sense that their acquisition does not require a process of exposure: they are primitives for the immune system. Other applications of the template—to language, concepts, the digestive system, or what have you—would be justified to the extent they cohered with successful scientific theorizing that posited primitives within a domain (including possibly, as the examples suggest, distinct applications to *sub*domains of the psychological).

We might alternatively characterize these primitives as what constitute the explanatorily relevant, normal *initial state* of some system of the organism. Such states need not be *temporally* initial: there could be elements that arise after a system's operations are engaged, but not *by* those operations, and then go on to serve as primitives for further operations. These elements could even be acquired as a result of a process that, while not part of the system, involves acquired elements of the system (again, we will return to the idea of acquired concepts triggered by learned stereotypes in Section 3.4.2 below). In the limiting case, the initial state would be the genotype itself (see above). But one need not suppose that the explanatorily interesting initial state for each such system—and thus what counts as innate for the system—is the same in each case or cannot include effects of the environment. In particular, on this view, the possibility arises again of one and the same trait being innate (primitive, initial) for one domain or system but not for another. It might be required, however, that no matter the domain or system, its initial state should not be itself acquired via learning, on pain of loosing contact with our pre-theoretic conceptions of innateness altogether. Such talk of domains, initial states, and proprietary processes (and their proprietary inputs) begs for clarification. But arguably they carry their weight to the extent that successful explanatory strategies find them useful.

Have we thus been led back to Canalization as cashed out negatively in terms of precluded environmental interaction? Recall the belief that one has a nose. (PRIM*) excludes this trait by dint of the process by which it is acquired in normal environments. (CAN) excludes it because it is not invariantly possessed in all *relevant* environments, where the nature of the trait and our explanatory project dictate that environments lacking experience necessary for acquiring this trait be deemed relevant (even if only conceptually possible). The basis for exclusion is thus differently characterized. But presumably, in the application of (CAN), the experience-deficient environment is deemed relevant precisely because such traits normally depend on acquiring the belief based on such experience; this is similar for other cases. What underwrites innateness for both (CAN) and generalized (PRIM*) thus appears to be the same: strategies that succeed at explaining the development of systems from their initial states by identifying proprietary environment-involving processes. Arguably, however, the nontemporal notion of initial state derived from Primitivism differs from the more naturally temporal notion common to the developmental biological models that inspire Canalization.

2.3. Suppose We Do Not Know What Innateness Is

We have discussed several conceptions of innateness, suggesting indeed that they may not differ as much as it can seem at first glance. But suppose that upon even closer inspection none of them pan out, so that we do not currently possess an explanatorily useful conception of innateness. Would this render the concept, and debates involving it, empty or unintelligible? Not necessarily.

First, it is a common semantic externalist claim about natural kind concepts that possessing them is compatible with holding many false beliefs concerning the kind—and indeed with possessing practically no conception at all (cf. Section 3.4). If this is right, then it is entirely possible that our talk of innateness has sufficiently “locked on” to a real phenomenon *even if* we are currently unable to characterize that phenomenon satisfactorily, as in the case of ordinary talk of, for example, weight, germ, and jade. It is perhaps even possible that we have locked onto multiple distinct phenomena in different areas of inquiry. That said, a concept's explanatory utility is reduced to the extent we lack an articulated conception, since a greater understanding facilitates the integration of explanations employing the concept with other explanations and theories.[14]

Second, some “innateness” debates could retain their significance *even if* there is no *phenomenon* of innateness that they concern. For example, the idea behind

[14] Several authors (Mameli and Bateson 2006; Samuels 2007) have recently suggested a strategy worth exploring on this front. Instead of searching for informative necessary and sufficient conditions for innateness, one might attempt to uncover empirical correlations among properties that serve as evidence for innateness—including among properties that have been mistakenly identified, singly or jointly, with innateness. Cf. Boyd (1991) on natural kinds and homeostatic property clusters.

Psychological Primitivism—that innateness debates in psychology concern what is learned—retains its interest, even if there is no unified phenomenon of what is *not* learned. The point is not just that the complement of a natural kind need not itself be a natural kind in the way that *mammals*, may be, but *nonmammals* surely is not. Rather, the point is that the complement in this case—the psychological primitives—need not contain any scientifically interesting *sub*-kind of the innate. What really matters is whether a trait is acquired through learning, not whether—if it was not learned—the process was in some sense normal or abnormal.

We find support for this second point both in the main arguments made on behalf of psychological nativism and in the main replies. Consider Chomsky's "poverty of the stimulus" arguments for our linguistic competence possessing an innate component. These arguments stress how children come to possess a linguistic competence that far outstrips the evidence available in their experience, and that consequently that competence could not have been acquired on the basis of that experience alone. Replies typically consist in calling attention to further evidence available to the child after all, or in arguing that learning strategies can extract from experience more information than was realized. It does not matter, so far as such debates are concerned, whether the alternative to learning is usefully labeled "innate," or that the label picks out a natural kind (even if it matters to what extent the details of the alternative acquisition story can be supplied in a given case). It is enough that there is an initial state without which the further perceptual input to the system would be inadequate to explain the final stable state of language acquisition that a normal human being achieves.

This places great weight on the question: what counts as learning? But perhaps that is as it should be, at least for some innateness debates. Of course, not all innateness debates can be rescued in this manner, since (as we noted even for psychological traits) not all center on learning. If there is no innateness phenomenon but these debates are to retain their significance, it will have to be for other reasons. But Psychological Primitivism's focus on learning suits particularly well the subject of our next section—the innateness of concepts. As we will see, much depends on what counts as rationally learning a concept versus merely having it triggered consequent to a perceptual process.

3. Innate Ideas

As mentioned in Section 1, it was nativism regarding *concepts* that originally divided traditional rationalists such as Leibniz and Descartes, who claimed most of our concepts were innate, and empiricists, such as Locke, Berkeley, and Hume, who claimed they were "derived from experience." Although Chomsky's proposals revived Rationalist views generally, Jerry Fodor's (1975) seminal book, *The Language of Thought*, revived the specific Rationalist views about *conceptual* nativism. In chapter 2 of that book, he proposed the following, radical conceptual nativist hypothesis:[15]

> (RCN) To a first approximation, all concepts expressed by single morphemes in English are innate.

(Morphemes are the smallest linguistic units having meaning; thus "do" and "un-" are (mono-)-morphemes—MMs–in English; "undo" is *poly*morphemic.) Standard estimates of the MMs used by a standard English speaker are between fifty thousand and 250,000 words (see Bloom 2000), and, of course, more serious approximations would have to take account of concepts expressed by MMs in one language but not in another (e.g., "chic"), or not yet in any.

An understandable reaction to (RCN) is to reject it as obviously absurd. Over 250,000 innate concepts?! But it is not clear what entitles someone to this reaction. What does anyone know about precisely what concepts are and how they are acquired; or exactly what "learning" is, and how or whether it should be contrasted with being innate? One merit of Fodor's audacious view is that it calls attention to the need to think about these and related issues with a lot more care than has been traditionally bestowed. In a useful discussion, Laurence and Margolis (2002, 26–27) call it "Fodor's Puzzle of Concept Acquisition," and rightly compare it to puzzles about induction raised by Goodman and about translation raised by Quine. For these reasons, we will organize the discussion around Fodor's view, even though in the end we will express some sympathy with its critics. We will first set out a number of issues that are crucial to the debate (Section 3.1), turning then to the main arguments Fodor presents in his 1975 and 1981 discussions (Section 3.2). Those arguments will lead us to consider the main internalist (Section 3.3) and externalist (Section 3.4) views about the nature of concepts, and the role of prototypes as internal "schemata" relating concepts and percepts. We will conclude (Section 3.5) with a brief discussion of Fodor's (1998, 2008) latest views, and of the processes of learning versus triggering that seem to lie at the heart of the dispute.

3.1. Preliminary Issues

3.1.1. What Concepts Are

We shall presume that concepts are the constituents of the objects of the so-called "(propositional) attitudes," such as *think* or *expect*. Thus, people who think fish dream, are thinking the proposition [Fish dream], which they can do only if they have the concepts [fish] and [dream] (we designate propositions and concepts by enclosing in square brackets the words that express them). For simplicity, we shall

[15] That book was seminal for a number of important views, another being that there is a language of thought, which serves as the vehicle of computation in the brain. *Pace* Churchland (1986, 389), (RCN) is entirely independent of this latter hypothesis.

also assume that concepts also serve as the meanings of words, and we will acquiesce in Fodor's treatment of concepts as mental representations.[16]

A crucial feature that concepts arguably need to possess to be effective constituents of attitudes is what Fodor (1998) calls "publicity":[17] it must be possible for different people, and the same person at different times, to have the *same* concepts. Someone cannot share a thought with someone else, or even remember a thought of her own, unless some of the constituents remain the same across people and time. This is a nontrivial requirement, since it is not at all obvious which facts about cognitive life are in fact stable across people, whose attitudes are constantly changing as they experience and think about the changing world around them. On Fodor's view, concepts are *type* representations individuatable in part by their content, tokens of which occur in different brains and in the same brain at different times.

3.1.2. Learning versus Triggering

Virtually all parties to the debate agree that experience plays a role in arousing whatever innate dispositions people may have to form concepts. But central to the debate between Rationalists and Empiricists is the question of just what the character of that role may be. Learning as popularly understood can include both rational and more brute-causal effects of perceptual experience, and this distinction is crucial to understanding the difference between Rationalists and Empiricists, neither of whom want to deny that experience plays *some* role in the causation of conceptual activity. Empiricists typically want to claim that most of our concepts are in some *rational* fashion *learned* on the basis of experience; Rationalists claim that experience merely serves to "occasion" the activation of a concept that a person already possesses. Descartes (1641b/1970, 227), for example, was especially impressed by the fact that the geometric figures studied by mathematics are not physically possible objects of sensory perception, and so thought that the concept [triangle] couldn't possibly be learned from experience. What we see are various irregular figures that in a context activate, or "occasion," the concept, causing us to see the figures *as* triangles. In an extreme form, the proposal is sometimes that the activating stimuli need not be related to the activated representation in any way that is rational or relevant in terms of intentional content: all that experience does is to provide stimuli that "triggers" the innate conceptual disposition, in the way that, say, an innate immune reaction is triggered by exposure to a certain pathogen, a reaction that is presumably neither rational nor intentional (see Jerne 1985). As

[16] Not everyone acquiesces here, but space forbids discussing Peacocke's (1992), Zalta's (2001) and Rey's (2005) preferences for treating concepts as more like Frege's (1892/1966) "senses." See Margolis and Laurence (2007) for discussion. We also abstract from the related issue of "nonconceptual" content (see Crane 1992).

[17] An unfortunate term, suggesting that concepts need be *social*, as opposed to "private," along lines claimed by some interpreters of Wittgenstein (1953, §§ 258ff), a debate entirely orthogonal to the present one. A better term for what Fodor has in mind is *stability* (see Rey 1983).

Fodor (1981, 304) expresses the Rationalist creed: "Simple concepts which arise as the effects of such triggering are, no doubt, learned *in consequence* of experience; experiences are—directly or indirectly—among their causes. But they aren't learned *from* experience."[18]

3.1.3. Evolutionary Worries

Another common reaction to (RCN) is to worry about how 250,000-plus specific MM concepts could possibly have evolved. Steven Pinker (2007), for example, finds it "hard to see how an innate grasp of carburetors and trombones could have been useful hundreds of thousands of years before they were invented" (95). But here, as so often elsewhere, it is crucial to distinguish *evolutionary* theory from the specific mechanism of *natural selection*. There is not the slightest doubt that humans evolved from earlier primates, but just which traits can be explained by processes of selection is increasingly controversial (see Sterelny and Griffiths 1999). Humans display abundant capacities (e.g., in music, science, logic, and mathematics) that far exceed anything that selection itself plausibly required.

Everything depends here upon just how one counts "traits" (*stomach? stomach and intestines? stomach and intestines and heart?*), as well as the character of the genetic options available at the time of selection. Thus, bilateral symmetry and the structure of bodily organs may reflect more about the underlying physiochemical structures, or "laws of form" (Thompson 1917), than anything about the specific demands of the selecting environment.

Moreover, animals might be evolutionarily prepared in specific ways to deal with items and states that may not have been present at the time of the evolution. To take the example that we have already mentioned, and to which many have compared (RCN), Niels Jerne won the Nobel Prize for his work arguing that all the antibodies in the immune system are in some sense available innately, waiting to be triggered by antigens. Similarly, even concepts of things that were not remotely available on the savannah might be innately available: after all, perhaps people could only *invent* things that counted as carburetors and trombones because they were innately endowed with the concepts of these things in the first place, just as many birds sing only their species' songs, or spiders weave only certain kinds of webs.

Of course, many will still find an individual innately possessing 250,000-plus concepts/songs/dispositions outlandish, and demand that the burden be on the conceptual nativist to provide a positive account about how there could be *so many*, and *such specific* ones. However, until we have a much deeper understanding of the structure of our minds and how it might be related to underlying

[18] Fodor's wording here, "learned in consequence of experience," seems to imply that he thinks concepts are learned after all. We take this to be something of a verbal slip (or issue), the crucial distinction remaining whether there is a *rational* relation between a concept and the experiences that cause it.

genetic structures, it is difficult to know how to count these things, and generally what constraints selection places on psychological theory. In any case, conceptual nativists have presented serious arguments for their view, to which we now turn.

3.2. Initial Arguments for Conceptual Nativism

Poverty of stimulus arguments (see Section 2.3 above) provide one serious argument for (RCN): experience simply does not seem to supply enough information for children to learn the concepts of mathematics, geometry, or enduring substances, or the host of word meanings that they acquire with astonishing speed and uniformity in their first several years (see Bloom 2000). The ur-argument that pretty clearly establishes this stimulus poverty is really due to a famous problem raised by Chomsky's teacher, the philosopher Nelson Goodman, regarding the concept [grue]. This is an artificial concept he devised with the meaning "applies to all things examined before *t* just in case green but to other things just in case they are blue" (Goodman 1954/1983, 74). Now, all the emeralds people have so far examined have been green, and human beings seem to happily generalize this property to emeralds they have not examined, including those that will be examined only after AD 3000. But all the emeralds they have examined so far have in fact also been grue—since, after all, they have been examined before AD 3000. Why do not people generalize to *that* property instead of green, so that, after AD 3000 they would be surprised to find emeralds *green*? All the sensory data we receive at least before AD 3000 will not decide the matter. What is worrisome here is that there is an *infinitude* of such bizarre concepts (one for each increment of time beyond AD 3000). Clearly, childhood stimuli are in principle too impoverished to decide between them. So at least in that respect, at least a sensory concept such as [green] can seem to be innate.[19]

[19] There is a vast literature on this "new riddle of induction," as Goodman (1954/1983) called it (see, e.g., Stalker 1994 for a representative collection). It is worth noting that, although Goodman himself did not see it as inviting a nativist moral, it is this very problem that was partly responsible for the resurgence of interest in conceptual nativism, since Chomsky (1955/1975, 33–34; 1971, 6–8) saw his ideas about the innate biases of children to only a specific subset of possible languages to be simply another (rather more elaborate) instance of Goodman's riddle. Indeed, given that an infinite number of grue-like concepts can be constructed for *every* concept we have, it is worth wondering why their existence *alone* does not establish that *none* of the concepts we naturally employ are learned! Fodor, however, though he does draw from Goodman's riddle nativist conclusions concerning the ordering of hypotheses (1975, 39), does not directly connect such considerations to conceptual nativism—perhaps because he suspects that these unnatural concepts are ruled out by some general considerations that enter into concept construction. But the burden would be on the defender of conceptual learning to say what these might be.

It is surely no news to empiricists that at least dispositions to sensory percepts (e.g. example experiences of green) need to be innate, and so perhaps the concepts (e.g. [green]) derived directly from them. After all, the mind cannot really be an entirely "blank slate," lest it not be able to learn anything at all. So it is often conceded that there are, as Quine (1969) put it, innate "quality spaces," biasing a child to generalize to [green] rather than [grue], and these would then serve as the basis for constructing all the concepts the child acquires. But how does the child do this? How does she "construct" new concepts on the basis of experience? This is where Fodor's (1975, 1981) arguments become germane (we will return to their bearing on the "grue" problem in Section 3.5).

Fodor's initial (1975, 79ff; 1981) argument for the innateness of concepts was quite simple. He pointed out that standard accounts of learning treat it as a process of *hypothesis confirmation*: in a classic discrimination experiment, for example, an animal learns to respond differently to Rs versus not Rs presumably by confirming an hypothesis along the lines of

(L) x is R iff x is red and triangular

and it does this by keeping track of instances of R that are, and those that are not, red triangles. But although this may be a perfectly clear way for the animal to acquire the *belief* that (L), it does not seem to be a way to acquire the *concepts of* [red] and [triangular], or even the conjunction of them: for in order to confirm (L) the animal must already have the means to *entertain* it, which it surely cannot do *without already possessing its constituent concepts*. At best, a red and triangular stimulus could *trigger* or *occasion* the already available concept [R], not *introduce* it.

It is important here to distinguish three notions of concept possession that are not clearly distinguished in most discussions:

(i) Having a symbol expressing a certain content that is actually *activated* in some mental process;
(ii) Having a symbol expressing a certain content that has never actually been so activated, but is *lying in wait*;

and

(iii) Being able to *express* a concept by logical construction on symbols expressing certain contents, either already activated or lying in wait.

The above argument of Fodor's (1975) establishes that, if learning is hypothesis confirmation, then an organism cannot learn any concepts whose contents it cannot already *express* in the sense of (iii). Thus, he allows the obvious point that not all concepts are actually *activated*. However, if learning is hypothesis testing, then it requires, in the very hypothesis, the activation of a concept either already activated or lying in wait. In the above example, the organism learning that something is R iff it is red and triangular must in that sense already have the concept [R].

This conclusion, however, might only disturb an Aristotelian, who thought that expressive power was actually acquired by properties in the world being *transmitted to* (indeed, being instantiated in!) the mind (cf. Section 1 above). A more modern empiricist might argue that what the organism *can* learn are syntactically novel ways to construct representations of the same conceptual contents it might always have been *able* to express, which it then sometimes *abbreviates* as MM concepts for use in memory and thought, as when we "chunk" material for use in short-term memory (see Miller 1956).

In his (1975) and (1981), Fodor supposes that we can and do learn new *composite* concepts by logical construction, which seems to suggest that he thinks an organism can increase its *expressive power* in this way. He seems to suppose, for example, that one may well increase one's repertoire of concepts by thinking "that thing is both red and triangular." (We will see in Section 3.5 that in his (2008) he sees this supposition as confused, and rejects even the learning of composite concepts if that means increasing expressive power.) All that he is concerned to deny in his (1975) and (1981) is that such constructions offer an account of *MM concepts*, which he thinks are *not* acquired by abbreviatory chunking; even when not activated, they must be lying in wait.[20] And he thinks this because he finds defective all arguments and efforts to establish that MM concepts are such abbreviations. Indeed, the crucial part of Fodor's (1975, 1981) view is his rejection of the classical empiricist account of the "analysis" of MM concepts, whose history has in fact not been a happy one, and which we now briefly review.

3.3. Internal Constructions

3.3.1. The Classical View

The Classical View of concepts treated them roughly as representations that are in some important way decomposable into conditions that are individually necessary and jointly sufficient for satisfaction of the concept, and are known to any competent user. The standard example is the especially simple one of [bachelor], which seems to be analyzable as [eligible unmarried male]. More interesting ones that served as the inspiration of much of "analytic" philosophy were Weierstrauss's analysis of the concept of a mathematical limit and Frege's (1884/1950) analysis of the concept of number.

[20] It is worth bearing in mind that one of Fodor's (1975) most important intended targets was Piaget (1954), whom Fodor read as claiming that expressive power could be increased by learning, as a child progressed through various cognitive stages. Whether Piaget needs to be read this way or could be read as intending only this more modest actual activation view is an issue to which we will return in Section 4.

Some of Fodor's own formulations can suggest that triggering involves, or can involve, the *acquisition* of a concept, not just the *activation* of a concept already possessed. Cf. Section 2.2.2 above.

3.3.2. *Empiricism and Verificationism*

The Classical View, however, has always had to face the difficulty of *primitive* concepts in which a process of definition must ultimately end. As mentioned earlier, seventeenth-century empiricism had a simple solution: all the primitive concepts were *sensory*. In the work of Locke, Berkeley, and Hume, this was often thought to mean that concepts were somehow *composed* of introspectible mental items ("images," "impressions") by *associations* among basic sensory parts. Thus, Hume (1734/1978) analyzed the concept of [material object] as involving certain regularities in our sensory experience, and [cause] as involving spatiotemporal contiguity and constant conjunction (see Elman et al. 1996 for recent associationist proposals along more neurophysiological lines).

Mere association, however, is not really adequate to capture the stability of concepts or the roles they play in thought. One person's sensory associations with [justice] may differ vastly from another's without the two people failing to share the concept, and few if any such associations (e.g., a blindfolded woman) are candidates for the analysis of such an abstract concept. Moreover, after Frege's work in logic, it became clear that any account of how a complex concept might be constructed out of sensory ones had to include an account of *logical structure*. This is precisely what many Logical Positivists attempted to provide. They focused on logically structured propositions instead of images and associations, and transformed the empiricist claim into the famous Verifiability Theory of Meaning: the meaning of a proposition is the means by which it is confirmed or refuted, ultimately by sensory experience; the content of a concept is the means by which experiences confirm or refute whether something satisfies it. The theory gave rise to sustained reductionist programs, such as those of *phenomenalism* (reducing material objects claims to claims about sense experience) and *analytical behaviorism* (reducing claims about the mind to claims about physical behavior). Note that it is no accident that empiricism has tended to be the view *both* that concepts are verification conditions *and* that they are acquired from experience, since verification conditions, for empiricists, are ultimately sensory tests.

Verificationism, however, came under much attack in philosophy in the 1950s and 1960s. In the first place, few, if any, successful analyses of ordinary concepts (like [material object], [expect], [know]) in purely sensory concepts have ever been achieved (see, e.g., Ayer 1934 for some proposals, and Quine 1953 and Putnam 1962b/1975, for criticisms). And this is not surprising. The relation of a concept to the sensory evidence for its application is generally quite complex, and dependent on indefinite numbers of other things being (believed to be) in place: someone's *looking* ill confirms their being ill only if (it is believed that) the lighting is right, one's eyes and brain are in working order, and there has been no deceptive mischief. Change the background beliefs and one may well change the relevant sensory evidence, but without necessarily changing the concept (say, of being ill). As Quine (1953, 41) famously put it, "our beliefs confront the tribunal of experience only as a corporate body" (a view called "confirmation holism"), which circumstance he saw as undermining any effort to supply such analyses, or "analytic" truths, as distinct from ordinary, "synthetic" beliefs about the world.[21]

Moreover, there is the difficulty of drawing a principled limit to how deviant people can be about the inferences they draw with concepts they seem nonetheless entirely competent to use. *Conceptions* of a phenomenon may vary wildly while a *concept* remains the same. Philosophers are notorious for defending outrageous claims, for example, that material objects are ideas, that rocks are conscious, or that contradictions can be tolerated. Such cognitive states present a serious prima facie difficulty for a theory of concepts that claims their identity or possession involves specific connections to other concepts or experiences.

Perhaps the most important argument for conceptual nativism is one that is implicit in Quine (1953), but has been increasingly explicit in the work of Chomsky (1968/2006), Harman (1964), and Lipton (2004), drawing upon Peirce (1903/1998, 287): the role of "*abduction*," or "inference to the best explanation" in both ordinary and scientific reasoning. Abduction is a form of nondeductive inference that, unlike *induction*, may involve the introduction of terms *not in the observational (especially sensory) vocabulary*. Thus, when physicists infer the existence of elementary particles and subatomic forces from ordinary macroscopic data, or even when a jury finds a defendant guilty on the basis of the evidence presented in a trial, they are often not merely making some sort of statistical generalization about that data, but leaping to the activation of what seem to be concepts such as [quark] or [corporate conspiracy] that manifestly involve commitments that go far beyond the concepts being deployed in characterizing the evidence.

In view of these issues about confirmation, it is unlikely that concepts in general could be defined in terms of the evidence adduced for them, and insofar as ordinary concepts are deployed to explain regularities in experience, it has seemed similarly unlikely to many philosophers that those concepts could be defined in terms of that experience.

There have been a number of responses to these problems with the Classical View, internalist and externalist. The internalist ones, to be discussed in the next three subsections, simply alter the Classical methods of internal mental "construction" of concepts and so continue to support Empiricism. It is because Fodor is equally sceptical of these further methods that he follows a number of philosophers in advocating an Externalist theory, to which we will in turn in Section 3.4, and which we will see affords a basis for his extreme Rationalism.

[21] A verificationist could go on to insist, as Quine (1960, 1969) is standardly read, that meaning would be similarly holistic ("meaning holism"). But, as Fodor and Lepore (1992) point out, this threatens to undermine the *stability* of concepts, given that no two people (or different stages of one person over time) are likely to share anything like the *totality* of their beliefs (only "threatens," since meaning holism—although perhaps not verificationism—can be defended in a variety of ways against this worry; see Greenberg and Harman (2006) and Pagin (2006)).

3.3.3. *Prototype Theories*

"Prototype" theories are sometimes taken to be a kind of definition theory (e.g., something is an F to the degree that it resembles a prototype), but more usually as a rejection of the demand for strict definitions, replacing them with descriptions of typical instances, or traits shared by typical instances. Rosch (1973) and Smith and Medin (1981/1999) showed that people respond differently (in terms of response time and other measures) to questions about whether, for example, penguins as opposed to robins are birds, in a fashion that suggests that concept membership is a matter not of possessing a Classical analysis, but of distance from a prototype or typical exemplars, robins being thereby better birds than are penguins.

Prototypes are clearly music to an Empiricist's ears, since constructing a prototype is an activity very much rooted in one's experience, and the character of the examples of a concept that one happens to have encountered. In her response to (RCN), Cowie (1999, 146–47) proposes that concepts do have, at least partly, a prototypical structure, and that it is this aspect of a concept that is learned from experience (see also Prinz (2002) on "proxytypes").

A number of objections have been raised against prototype theories of concepts:

(i) Not all concepts have prototypes (consider [carburetor] or [not a cat]), and, even in the case of those that do, it just does not seem plausible to insist that a competent user of a concept be acquainted with one, much less the *same* one for everyone: Spaniards and Australians may have different prototypes of [bird] but nonetheless share the concept;

(ii) Many concepts cannot be identified with prototypes, since competent users of a concept know full well there can be perfectly good instances of a concept (e.g. [bird]) that are not prototypical (e.g., penguins), and, moreover, that something could satisfy the prototype (feathered, chirps) without being an instance (e.g., fancy toys);

(iii) In order for prototypes to figure effectively as an account of concepts, some "distance" metric among them would need to be specified (how unlike a typical bird can something be and still be a bird?), and it is not clear how this is provided without an independent characterization of the concept;

(iv) Prototypes are not *compositional*: (knowing) the prototype of [pet] (e.g., a dog) and the prototype of [fish] (e.g., a trout) does not determine (knowing) the prototype of [pet fish], and someone might know the prototype of [pet fish] without knowing the prototypes of [pet] or [fish] (see Fodor 1998).

Prototypes arguably play a role in how people quickly *tell* whether something satisfies a concept, that is, they provide good *evidence* for its application, and may figure centrally in a person's *conception* of its extension, but, as we saw in rejecting verificationism, evidence and conceptions are one thing, conceptual content quite another (see Rey 1983/1999, 1986; and Fodor 1998, 2001 for further discussion).

3.3.4. Non-Sensory Primitives

There is no intrinsic reason for either the Classical or maybe even Prototype views to be committed to either a sensory or other verificationist theory of concept construction, and there have been both philosophers and cognitive scientists who have thought that certain very general concepts, such as [agent], [object], [property], [number], [cause]—sometimes called "framework concepts"—might be primitive and innate, and provide a better basis than mere sensory experience for conceptual construction (see Miller and Johnson-Laird 1976; Moravcsik 1975; Pustejovsky 1995; Jackendoff 1983, 1992; Pinker 2007, ch. 3, for numerous proposals along these lines, and Fodor 1998, chs. 3–4 for criticisms).

A problem for this approach is to specify the relevant framework concepts and, more importantly, provide an account of what (constitutively) determines their content. Proponents will maintain that this is settled simply by whatever primitives their successful explanatory programs posit, but it is not clear how those programs as they are currently pursued will suffice. Fodor (1998, 49ff) complains that in order for such explanations to go through, the framework concepts must be understood univocally, and this univocality needs to be established. It is not enough merely to appeal to homophonic English words. Thus, Fodor argues, Jackendoff's (1992, 37–39) example of [keep] seems on the face of it either ambiguous or polysemous in "keeping a bird," "keeping time," "keeping a crowd happy," and "keeping an appointment." Perhaps it is, if [keep] can be analyzed into some complex such as [cause a state that endures over time]—that is, if [keep] is not a framework concept after all. But now we have to ask the same question about *its* constituents: is [cause] or [endure] univocal? Does an appointment *endure* in the same sense as money does? Are they both *caused* in the same sense? What sense is that?

A related general problem with any approach that looks to analyses of any sort is to provide a basis for claiming that some proposed construction really is the *correct* account or "analysis" of a concept, as opposed to simply some banal or deeply entrenched belief about the world. Is it part of the concept [cat] that cats are animals? Or is this merely a belief that everyone takes for granted? Would thinking that cats are robots controlled from Mars be as incoherent as thinking that there are cats that are not cats? This is a form of the challenge already mentioned that Quine (1953, 1956) raised against the analytic/synthetic distinction, to which no generally accepted reply has yet been made (see Putnam, 1962a/1975; Katz 1990, 216ff; but also Horwich 1998, 2005; and Rey 2009, for recent proposals to meet this challenge that are related to Fodor's own proposals).

3.3.5. Other Methods of Definition

It is notable that many of Fodor's (1981) examples of constructions are *Boolean* (composed of simple truth functions such as [and] and [or]). But there is no reason to suppose that mental constructions might not be more sophisticated, employing complex quantifications and modal and probability operators. For example, concepts such as [energy], [mass], [force], and [space] are likely best defined *together* by setting out the complex laws of physics in which they all occur, and adding "and that's all there is to being any one of these things." A raft of concepts is thereby introduced in terms of the roles they play together in explaining some domain. The philosopher Frank Ramsey (1929) developed a technical proposal, now called "ramsification," that allows this to be done with a great deal of precision (roughly, a "Ramsey sentence" says that there does indeed exist a number of things, say, energy, mass, etc. that satisfy the terms that a theory introduces with the conjunction of all its claims; each of the individual terms can then be defined by its role with the others in that long conjunction; see David Lewis (1972/1980) for a lucid exposition). Such clauses in a theory might serve for what Carnap (1952) called "meaning postulates," or principles that are set out as defining of the terms being introduced by them (Murphy and Medin, 1985/1999; Block 1986). This would seem to be in part what Susan Carey (2009, chs. 8 and 11) has in mind in defending a "bootstrapping" proposal she finds in Quine (1960, 1969) and other philosophers of science. According to her proposal, "place holder" symbols are generated by assimilating a lot of information about a domain in the form of diagrams, lists, stray claims, and sometimes some serious theory: one could think of all this material as being expressed by one long Ramsey sentence that conjoins all of it, and allows one to refer to the single phenomenon (if any) in the world that satisfies it. The gradual assimilation of this material permits a learner to gradually master new concepts by slowly grasping partially now one conjunct and now another until they all fit together to characterize a stable concept.

One prima facie problem with this approach, however, is still the above Quinean one of deciding which clauses in a theory are to be included as meaning constitutive, and which as merely empirical claims about the world. Moreover, people are constantly creating and revising new theories involving old concepts, without ipso facto changing the content of those concepts. When Darwin proposed that humans are a kind of primate, he may have changed the prevailing *conception* (i.e., common beliefs) about human beings, but he did not thereby change the *content* of the concept. Indeed, it was precisely because he was employing a concept, [human], with the very same content as creationists that the latter were so upset!

More modestly, however, ramsifications might serve as "reference fixers," or descriptions that *in a particular context* serve to fix the reference of a term without being synonymous with it (see Jackson, 1998). This idea has its origins in Kripke's (1972/1980) discussion of proper names, where he points out that the descriptions people commonly associate with a proper name (e.g., "the discoverer of America" with "Columbus," "the author of *Moby Dick*" with "Melville") are certainly not synonymous with the name, and, indeed, could turn out to be false of its referent. He argued that proper names and many natural kind terms are "rigid designators" that name the same thing in all possible worlds, including those in which the common descriptions are not necessarily true (for example, it was perfectly possible for

Columbus not to discover America): the common descriptions serve merely to "fix the reference" of a term in a particular context, and not as definitions (an idea to which we will return below in considering Laurence and Margolis's (2002) "kind syndromes").

Kripke's proposal caused a minor revolution in philosophy, for it reinforced a suspicion that had been sown by Quine's attack on the analytic that *no* internal, *epistemic* condition would suffice for a theory of meaning. So, again, the Classical empiricist view of concept acquisition seemed implausible. Rather, meaning must in the first instance essentially involve some sort of relation to phenomena *external* to the mind and brain of a thinker.

3.4. Externalism

3.4.1. Causal Theories

Kripke's idea that reference is constituted at least in part by external facts was already suggested by some of the views of the later Wittgenstein (1953), and was independently proposed by Hilary Putnam (1975). Tyler Burge (1979) applied it not only to the meanings of both natural ("water") and artifactual ("sofa") kind terms, but also to their corresponding concepts. They pointed out that such terms and concepts do not standardly involve definitions known to their users, but rather (causal) relations with their actual referents and/or the social community in which they are used.

Now, of course, if possessing a concept *did* depend upon having a certain historical connection to an environment, certainly a certain facile form of empiricism would be vindicated: what concepts one had would depend directly upon what world one had experienced. However, while these causal intuitions provide an interesting challenge to the Classical View, they are quite inadequate as a theory of conceptual competence. Mere causal interaction in a certain community and environment cannot be enough, since surely not *all* sentient beings in New York have the concepts of a Columbia University physicist they have bumped into on the subway! But if mere causation is not enough, and definitions are not available, what does determine whether someone has a specific concept?

A number of writers have proposed varieties of *counterfactual* causal links: x has the concept [y] iff some state of x did/would causally co-vary with the worldly phenomenon y; for example, x did/would *discriminate* instances of y under certain (ideal, normal, evolutionarily significant) conditions, as a matter of nomological necessity. Thus, someone has the concept [horse] iff she could under certain conditions tell the horses from the non-horses. This is the idea behind "informational" (or "co-variational") theories of the sort proposed by Dretske (1980, 1987), and Stalnaker (1984). Fodor (1991, 1998) developed the most sophisticated of these accounts with his "asymmetric dependency" analysis, whereby a symbol x means that p iff its application to non-p cases depends on its application to p cases, but not vice versa (thus, one calls a misperceived cat "a dog" only because one calls a dog "a dog," but one does not call a dog "a dog" because one calls a cat one). In such cases, Fodor (1998) talks of symbols "locking" onto their referents. (Notice that if the causal connection between a concept and its meaning-constitutive referent were *counterfactual*, then the above facile argument for empiricism would no longer be available: the causal relation would not depend upon *actual* history, much less experience.)

There are numerous problems with informational theories. Once one departs from the relatively straightforward cases of perceptual concepts that even empiricists would be happy to regard as primitive, it is by no means obvious that there really are the kinds of genuine laws linking the brain symbols to worldly phenomena in a way that would provide those symbols with their conceptual content (what are the laws for [cause]? [object]? [space]?, or for logical concepts such as [not] or [if]?) (see Loewer 1996). Particularly troublesome cases are "empty" concepts, such as [ghost], [angel], [soul], and (at least for some) [triangle] and [square], for which there are arguably no *possible* worldly phenomena with which brain symbols could co-vary.[22] Fodor (1998) crucially claims "there can be no *primitive* concept without a corresponding property for it to lock to" (165), but then it would appear that he would have to allow such empty concepts to be "constructed" after all. But, if them, why not others?[23]

These and other considerations have led a number of philosophers and psychologists to embrace "two factor" theories that claim that conceptual content is determined by *both* causal relations to the world *and* internal computational roles (see Block 1986; Carey 2009, 514ff). For example, the content of [bird] may consist both of a role component that figures in accounts of internal psychological processing—for example, common conceptions and prototypes associated with birds—*and* an external causal component that determines what phenomena in the world is picked out. However, many of the problems we mentioned with respect to these components considered separately would appear to persist in their combination: it seems possible for prototypes and conceptions to vary across people without variation in their concepts, and it is not clear how internal factors can figure in a compositional semantics.

In his (1998), Fodor proposes an interesting intermediate view that, while not treating prototypes as *constitutive* of concepts, treats them as playing a crucial role in the *triggering* of them.

3.4.2. Fodor's (1998) Prototype Triggering View

Fodor (1998, 127ff) considers an important problem for his view that concepts are triggered and not learned, what he calls the "doorknob/DOORKNOB" problem: why is it that [doorknob] is regularly triggered by doorknobs, and not, say, by

[22] Such cases motivate an appeal to Fregean senses, cf. fn 12.

[23] One might reply here that in the case of empty terms, such as "elf," "phlogiston." or "Vulcan," there is, indeed, nowhere to retreat but to the claims by which they are introduced, but that in the cases where terms *do* succeed in referring, the introductory material gets trumped by the referent, so that, e.g., "water" or "polio" is no longer tied to it (see Rey 2005 for discussion).

giraffes or French flags (see also Sterelny, 1989)? The fact that it is suggests that a concept perhaps is not merely "triggered" as a brute causal process, but, in Descartes's phrase, is "occasioned" by virtue of some kind of rational, confirmatory relation between a concept and perceptual stimuli, and so might involve a kind of learning after all.

To solve this problem, Fodor adopts a quasi-Kantian metaphysics of the very items so picked out. He exploits the prototype views we mentioned earlier, *not* toward making some point about the content of *concepts*, but instead about the metaphysics of, for example, *doorknobs*! The reason doorknobs occasion activation of the [doorknob] concept is that *doorknobs just are the kind of thing to which humans generalize when presented with prototypical doorknobs*. That is, he treats such cases on the model of the "response-dependent" properties that many philosophers have proposed for secondary properties such as color. (Unlike Kant, however, Fodor does not claim that *all* concepts are of response-dependent properties; in the next section we will return to the "scientific" ones that are not.) Whether such a story can really be told generally in conjunction with an informational theory of content without circularity remains controversial (the proposal skirts perilously close to the claim that possessing the concept of doorknob requires a representation being locked onto—well, just those things to which the representation is locked! See Cowie 1999, 96–99). However, Fodor (1998, 138) argues that circularity is avoided by the fact that at least many prototypes are known to be specifiable independently of the concepts they occasion (as in Rosch 1973).

Whether or not his view avoids circularity, Fodor sees it as permitting him to retreat from his initial (1975, 1981) controversial view that all MM *concepts* are innate. His 1975 argument too quickly assumed that if a concept was not learned, it was innate. Recalling the distinction between rational and brute-causal effects of perceptual experience, on the present view concepts can be *acquired* from experience by a nonrational triggering mechanism without being learned, and so need not be either learned *or* innate. As he puts it in his still later 2008 book:

> You can't infer from a concept's not being learned to its being innate; not, at least, if "innate" means something like "not acquired in consequence of experience." There would appear to be plenty of ethological precedents—from "imprinting" to "parameter setting" [in Chomskyan linguistic theories] inclusive—where it's implausible that the acquisition of a concept is mediated by a rational process like inductive inference, but where concept acquisition is nevertheless highly sensitive to the character of the creature's experience. (2008, 144–45)

This leads him to refine his 1975 position:

> What's learned (not just acquired) are stereotypes (statistical representations of experience). What's innate is the disposition to grasp such and such a concept (i.e., to lock to such and such a property) in consequence of having learned such and such a stereotype. (2008, 162)

Indeed: "the kind of nativism about [doorknob] that an informational atomist has to put up with is perhaps not one of concepts but of *mechanisms*" (1998, 142), the processes of which he takes to be nonrational and even nonintentional. They are simply "brute causal," not psychological processes.[24]

Slightly revising the distinctions we drew earlier between activation, lying in wait, and expressive power, we might put Fodor's (1998) view as follows: concepts themselves are not innate and do not lie in wait, unactivated. What is innate are highly specific dispositional mechanisms to acquire a concept upon exposure to the concept's prototype. These *mechanisms* lie in wait in a creature's brain, and the actual acquisition of a concept consists in a mental representation being produced by the activation of the mechanism, and coming to stand in a certain counterfactually specified locking relation to the real world phenomenon that constitutes its content. The notion of a conceptual system's *expressive power* is expanded from merely concepts activated or lying in wait to the range of concepts these innate mechanisms can produce in this way, and by logical combination. Fodor's main claim remains that expressive power in this sense is not increased by learning. The relation of a prototype to the concept it activates through these mechanisms is brute causal.

Incidentally, Fodor was not the first to despair of an intentional solution to relating percepts to concepts. In his *Critique of Pure Reason*, Kant (1787/1934, A138ff) postulated his "schemata" to mediate between the two,[25] and resignedly claimed that

> the schematism of our understanding, in its application to appearances and their mere form, is an art concealed in the depths of the human soul, whose real modes of activity nature is hardly likely ever to allow us to discover. (Kant, 1787/1934, A141)

For Fodor (1998, 2008), it is stereotypes that play the role of Kant's schemata, triggering innate dispositions to lock onto a property. But, as we saw above (Section 3.3.3), they do not exhaust a typical concept that has commitments far beyond them. So there is still the question why a specific stereotype triggers one concept rather than another, for example, [green] rather than [grue]. This is what Fodor (2008) thinks is determined by just brute causation.

Note that, *pace* Cowie (1999, 72), Fodor's despair here does not imply any "non-naturalism" on his part. Fodor's view is not that there is no *natural* explanation of acquisition; there is just no *intentional* one. Explanation of *most* of nature is not intentional. Why should it be surprising if concept acquisition turns out not to be so?

[24] This emphasis on whether the processes are brute causal or psychological recalls Samuels's conception of innateness (Section 2.2). But, of course, on that conception the concepts would therefore be innate as well, which Fodor here is allowing that they "perhaps" are not. (Cf. fn. 11.) It is unclear what of importance would be lost if Fodor embraced Samuels's view and continued to claim that concepts are innate as well.

[25] Unlike Fodor, Kant was here mainly concerned only with what he called "pure" concepts, such as [object] and [cause], and not with how to acquire a concept from experience, but how to apply one to it. And, of course, Kant's project is to determine transcendentally what is constitutive of experience, not to provide a naturalistic account of it. But the similarity in problems is striking.

Still, one might find such a view puzzling and unsatisfying: there remain those 250,000-plus *specific dispositions* to acquire concepts by specifically linked stereotypes, and this might seem even more profligate than merely 250,000-plus concepts by themselves. And so one might well want to continue to investigate further ways concepts could be constructed from experience without merely triggering specific innate dispositions. Fodor (2008) deals with this response.

3.5. Fodor's (2008) View

In his most recent discussion of the issues, Fodor (2008) provides a further argument against concept learning to those already noted above, one based simply on the familiar problem of distinct coextensional concepts, for example [morning star]/[evening star], [renate/cordate], [triangle]/[trilateral] (cf. Frege 1892/1966). How, Fodor wonders, can experience by itself provide a basis for learning the one concept and not the other? Only, he replies, by "representing the experience in different ways" (2008, 133–35), which again requires that one already has the very concept that experience has been recruited to teach.

Moreover, he scorns his earlier views as being "too modest" with regard to complex representations:

> What I should have said is that it's true and *a priori* that the whole notion of concept learning is per se confused . . . *no* concept can be learned, primitive or complex. (2008, 130 and 138)

Certainly, continuing the line of his (1975), construction of complex representations is not a way of acquiring an *expressive capacity* that one's nervous system did not already possess (we will return shortly to weaker, mere activation notions of concept possession, with which he may have been confusing expressive capacity in his (1975)).

Fodor (2008) also replies to some plausible proposals advanced by Margolis (1998) and Laurence and Margolis (2002), who argue that many ordinary natural kind concepts are acquired by children rather in the way that Fodor (1998) reserves for "scientific" ones. As we mentioned in the last section, in his (1998, 150–62) Fodor, unlike Kant, does allow that not all concepts are of mind-dependent phenomena. In particular, he allows that natural kind concepts, for example, [water], [gold], apply to things in the world that, unlike doorknobs, enter into laws, independently of what humans generalize to. Fodor claims, however, that these are "a late and sophisticated achievement," attained only "in the context of the scientific enterprise" (1998, 159). Although children and other proto-scientific people do in fact have such concepts, Fodor does not think they play a scientific role in their thought—they are for such people concepts of natural kinds but are not yet treated by them *as* natural kind concepts.

It is hard to see, however, why the intellectual attitudes (if not all the skills) of scientists are not *sometimes* available to nonscientists. In the work mentioned earlier on proper names and natural kind terms, Kripke (1972/1980) and Putnam (1975) drew attention to what would seem to be a property of much of the *ordinary* use of natural kind terms, such as "water" and "gold," whereby people are prepared to take such terms to refer to things with a "hidden essence" precisely along the lines Fodor reserves for scientists. Lest one suppose that perhaps Kripke and Putnam are securing agreement about these terms only from scientifically minded philosophers, it is worth noting that a number of cognitive scientists have produced evidence of similar intuitions in children and non-Westerners (see Macnamara 1986; Keil 1989; Medin and Atran 1999; Gelman 2003).[26]

In relation to concepts, Laurence and Margolis (2002, 38) develop a suggestion of Putnam (1975) and propose a quite general function (they call it a "kind syndrome . . . a collection of properties that is highly indicative of a kind yet is accessible in perceptual encounters. . . . [e.g.,] typical shape, motions, markings, sounds, colors, etc.") that, in conjunction with the essentialist attitude, takes an eliciting description to the nearest, contextually natural or explanatory kind that includes the material in the syndrome. Thus, [water] might be constructed as the nearest natural kind that includes the odorless, tasteless stuff that we drink, is found in streams and ponds, and is the stuff of rain. One might well extend this move beyond merely natural kind concepts to whatever is the "best account" of the reference of a term: thus, a nonnatural kind concept, [viol], might be constructed as the nearest kind that plays an explanatory role in the history of the modern violin (see Rey 1983). Carey (2009, 519) tries to generalize this account with her above-mentioned bootstrapping proposal, which she argues is not restricted to or dependent upon an antecedent notion of natural kind. Although such a kind syndrome is not *constitutive* of the concepts of, say, [water] or [viol], plausibly the only way someone could lock on to the corresponding properties is by knowing such clauses. It is, as Margolis (1998) puts it, a "sustaining mechanism" that explains the locking and keeps it in place. So, why not suppose that learning those clauses provides an account of ordinary folk *learning* a concept, even if it is not defining of it?

Fodor (2008, 144) notes in reply that "'You can learn (and not merely acquire) A' and 'Learning A is sufficient for acquiring B' just doesn't imply 'You can learn B.'" Again, B may merely be *triggered* by the learning of A. Remember, as we noted in

[26] Recently, Machery et al. (2004) have found evidence of intuitions purportedly contrary to those of Kripke and Putnam among some non-Western peoples. Such findings tend, however, to be beside the point, since what is at issue is not whether *all* uses of such terms are *always* to a hidden essence, but only that a significant class are so for at least some people other than scientifically minded philosophers, which this research in fact confirms: many of the non-Westerners shared the Kripke/Putnam intuitions. For further criticisms, see Devitt (2011).

Fodor (1998, 154–55) argues that the evidence that children have natural kind concepts as such is all interpretable merely as evidence that they make an appearance/reality distinction: the evidence of which he is aware does not establish that they think of a hidden essence as being the *cause* of appearances. This latter presumption is questionable: it is not at all clear that hidden essences must be (believed to be) causal in this way; many appearances, (e.g., the color of the sea) might be (believed to be) due entirely to factors other than water being H2o. But in any case, in reply to Fodor, Gelman (2003, ch. 5, esp. 135) provides just such evidence of causal thinking.

discussing abduction, the concepts we deploy in experience often involve commitments far beyond the evidence they explain, and so the question is just what takes us from the evidence to *these* and *not other* further commitments.[27]

A way to bring out the gap that Fodor is positing between the kind syndrome and conceptual locking is to consider the problem of saying just what the "nearest" kind is: point to a cat and you point *inter alia* to, say, its ear, a *flea* on its ear, to a *domestic cat*, to a *feline,* to a *mammal*, and so on. It is here that Goodman's "grue" problem we discussed in Section 3.2 returns with a vengeance. There are an infinite number of possible concepts that are compatible with any finite set of data a creature will have encountered: not merely an infinitude of grue predicates, in which "observed before *t*" is incremented by one year, but an infinitude of other analogous combinations, for example, [gred], [grellow], [catiraffe] (a cat and observed before *t* or otherwise a giraffe), etc. (Note that we cannot rely on human beings always referring to genuine natural kinds: consider [race], [humors], [earth], [air], [angel], even [red] and [green] themselves, which depend on idiosyncracies of our visual system.) Devitt and Sterelny (1987/1999) call this general problem the "*qua* problem": *qua*, or as, *what* do you refer when you point to the cat? Spelling out "nearest natural kind"—perhaps in terms of contextually relevant contrasts—would be a difficult, but perhaps not entirely hopeless research program. Fodor, though, is not holding his breath. He thinks the answer lies not in some rational connection between the relevant concept and the stimuli (even in the context), but in a brute-causal fact by which the concept is simply perceptually triggered by these stimuli.

4. Conclusion

Proposals such as Laurence and Margolis's (2002) proposal seem to come as close as one can to capturing the pre-theoretic notion of "learning a concept."According to them, at least with natural kind concepts, new primitive concepts can be acquired by learning, because there can be cases where one's experience "initiates a process where information is collected, stored, and manipulated in a way that controls a representation so that it tracks" what it is in fact about (2002, 43). On their view, because such concepts are learned, they are not innate. Fodor, on the other hand, would insist that only the *collected information* is learned, denying that the concept associated with it is therefore learned. The disagreement here can be seen as one concerning whether brute perceptual triggering is required in addition to the

[27] An interesting possibility raised by Fodor's view here is that someone might create a Ramsey sentence that *does not* trigger an innate (disposition to acquire a) concept. Perhaps this is the situation we find ourselves in with respect to, for example, the phenomenon of *light* as described in modern physics, which behaves both like a wave *and* like a stream of particles, while it is difficult for (most of) us to conceive how something could do both.

rationally assimilated information, but also in part as one concerning what should count as "learning," or at least learning a concept: can *it* include mere brute triggering by collected rationally relevant information or not?

Put aside the pre-theoretic notion of learning, as well as Fodor's last proposal to retreat merely to dispositions to acquire concepts upon exposure to prototypes. Remembering the distinctions we originally drew in Section 3.2 between *expressive power, lying in wait*, and *actual activation*, a boringly (or maybe radically) ecumenical view suggests itself. Some concepts could be *both* innate and learned at least in the following sense: concepts would be *innate* insofar as they cannot be constructed from or defined in terms of experience, but *learned* insofar as they are activated in an effort to confirm their application ultimately on the basis of experience.[28]

What is given at birth, or as a consequence of brute non-intentional influences on the brain, is a system with a certain expressive power: a set of primitives that have their content by virtue of their causal relations to the world, and principles of logical combination that permit the construction of an indefinite variety of logically complex representations out of them. These primitives and their logical combinations determine what the creature can *possibly* think, and this, if Fodor is right, can never be increased by any rational process.[29]

However, that the *activation* or *deployment* of a concept can often depend upon learning about the world, laboriously constructing stereotypes or Ramsey sentences that are designed to pick out the nearest explanatory kinds that offer, via abduction, the best explanation of relevant phenomena. The process could proceed much along the lines of a "hypothetical deductive" (HD) model of explanation, according to which explanation consists in deducing descriptions of the target phenomena from general hypotheses that are thereby confirmed (see Hempel 1965). Such a process is not the less confirmatory should the hypotheses, or their constituent concepts themselves, not have their *source* in experience, or not be reducible to it. (Of course, not all activation is by learning: some may be brute, and some may, as Hume (1734/1978, 10) emphasized, be merely imaginative—and here, indeed, one wonders what the source of imagined concepts could be other than the native repertoire, if they are neither triggered nor constructed from experience.)

There are, of course, constraints on how easily people actually can perform the logical constructions. To invoke a nice distinction from Chomsky (1965), there may be "performance" issues (e.g., short-term memory resources, motivation) constraining the activation of an underlying conceptual "competence." It is relatively easy to construct [viol] and [prime number]; much harder, [curved space-time]. The difficulties may even fall into patterns: construction of representations of

[28] The data to be explained in learning a concept in this way need not be evidence about whether a concept actually *applies to the world*. Often we come to grasp concepts by explaining merely the remarks of another speaker, as when we understand the concepts, say, [karma] or [phlogiston], of some theory we presume to be false.

[29] Note that this ecumenical view could be adapted also to the expanded notion of expressive power introduced in Section 3.4.2.

abstract concepts beyond perceptual appearances seems to require sometimes enormous effort (as in the case of the physics concepts of *heat* and *entropy*); and perhaps the difficulties are subject to developmental stages of the sort suggested by Piaget (1954, 1980). Indeed, even if concepts themselves (and not merely the dispositions to acquire them) were innate, Piaget could be right about childhood inabilities, say, to think about rational numbers. A neonate may well have a system capable of expressing every concept she will ever learn, but simply lack the further memory capacities or motivation to activate them.

The idea that one and the same trait could be both innate and learned might seem counterintuitive, if not downright paradoxical (cf. fn. 2). But it will perhaps seem less so when one considers the grounds for thinking of concepts in this way. The suggestion is that it is the full repertoire of concepts that is innate, with learning serving to select from the set of concepts those that are suitable for explaining the relevant data of experience. On this view, we thus must distinguish two levels of possession: concepts are innately had, but can also be had in a further, activated sense as a result of learning. Learning a concept thus turns out to be a kind of acquisition that involves an innate component—the concept itself—just as, if Jerne is right, acquiring certain immunities involves being exposed to certain pathogens that activate innate antibodies. Our ordinary, intuitive notion of innateness would simply turn out to pick out a phenomenon that is more intricate that it initially appears to be.

Just how much this conception of learning ought to satisfy conceptual empiricists depends upon both the degree to which concept activation *does* involve something such as an HD model of explanation of purely sensory material and the degree to which the HD model is an adequate model of learning. Neither of these questions can remotely be regarded as settled. If the locking of a representation onto an external content were caused merely by some purely "accidental" reference fixer—for example, a child who happens to lock onto quarks as a result of learning the description "what father wrote his book about" and using it to explain his father's scholarly behavior—it would seem pretty far from counting as *learning*.[30] But is locking as a result of mastering some facts in a textbook in principle any different? Even if they are not definitions, some stereotypes, reference fixers, and Ramsey sentences seem more "rationally connected" than others to the concepts they may trigger, but it is not at all clear how to specify just what that rational connection might be. Indeed, what *constrains* the further commitments a concept involves beyond the evidence that triggered it? Why is it—if it is—more rational for a child to jump to the concept [green] and not [grue] in encountering grass and pistachio nuts?

Conceptual nativism thus turns on a number of issues that may not have been evident at the outset:

[30] Consider a child who might be genetically incapable of understanding quantum physics: on some views, he might be said to think "his father works on quarks" simply by virtue of "reference borrowing" among members of a language group (cf. Kripke, 1972/80), but without his actually having *learned* the concept.

(i) whether empiricists can succeed in providing successful *analyses* of some significant portion of MM concepts from some set of primitives, despite their persistent failures to do so (see Section 3);
(ii) whether some sort of externalist, "informational" theory of content can be sustained on behalf of concepts lacking internalistic analyses (Section 3.4.1);
(iii) whether and the degree to which concept acquisition involves an HD or other "rational" process (and whether that process counts as "learning"), or, instead, a brute-causal, nonrational mere "triggering" process, lost in the mysteries of neurophysiology (Section 3.5);

and

(iv) whether there is any independent justification for supposing some 250,000-plus concepts, or specific dispositions to acquire them, are innate.

At this point, it would be foolhardy to suppose that any of these issues are anywhere near being settled. The possibilities of information theories of content and various means of concept construction have yet to be adequately explored, and, as we have emphasized, the psychological versus non-psychological processes of occasioning versus triggering a concept on which we have seen the question of the innateness of concepts seems to turn, have yet to be understood.[31]

REFERENCES

Adams, R. (1975). Where Do Our Ideas Come From? In S. Stich (ed)., *Innate Ideas*. Berkeley: University of California Press.

Aquinas, T. (1266/1948). *Summa Theologica*, translated by Fathers of the English Dominican. Province: Benziger Bros.

Ariew, A. (1996). Innateness and canalization. *Philosophy of Science* 63: S19–S27.

———. (1999). Innateness Is Canalization. In V. Hardcastle (ed.), *Where Biology Meets Psychology*. Cambridge, MA: MIT Press.

———. (2007). Innateness. In M. Matthen and C. Stephens (eds.), *Handbook of the Philosophy of Biology*. Amsterdam: Elsevier.

Aristotle (1968). *On the Soul* (*De Anima*), Books II–III, translated by D.W. Hamlyn. Clarendon Aristotle Series, Oxford: Oxford University Press.

Ayer, A.J. (1934). *Language, Truth and Logic*. Mineola, NY: Dover.

Bateson, P. (2000). Taking the Stink Out of Instinct. In H. Rose and S. Rose (eds.), *Alas, Poor Darwin*. London: Jonathan Cape.

[31] For helpful discussion, we are grateful to Jonathan Adler, Greg Ball, Paul Bloom, John Collins, Jerry Fodor, Eckart Forster, Barbara Landau, Joseph Levine, Edouard Machery, Yitzhak Melamed, Eric Margolis, Richard Samuels, and participants in a Philosophy of Linguistics Workshop at the University of Maryland.

Block, N. (1986), Advertisement for a semantics for psychology. In P. French, T. Euhling, and H. Wettstein (eds.), *Studies in the Philosophy of Mind*, vol. 10 of *Midwest Studies in Philosophy*. Minneapolis: University of Minnesota Press.

Block, N. (1995). How heritability misleads about race. *Cognition* 56: 99–120.

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

Bouchard, T., Lykken, D., McGue, M., Segal, N., and Tellegen, A. (1990), Sources of human psychological differences: The Minnisota study of twins reared apart. *Science* 250: 223–28.

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies* 61: 127–148.

Burge, T. (1979), Individualism and the mental. *Midwest Studies in Philosophy* 4: 73–121.

Carey, S. (1991), Knowledge, Acquisition: Enrichment or Conceptual Change? Reprinted in Margolis, E., and Laurence, S. (1999). *Concepts: Core Readings*. Cambridge, MA: MIT Press.

———. (2009), *The Origin of Concepts*. Oxford: Oxford University Press.

Carnap, R. (1952), Meaning postulates. *Philosophical Studies* 3: 65–73.

Carruthers, P., Laurence, S., and Stich, S. (eds.) (2005). *The Innate Mind: Volume 1: Structure and Contents*. Oxford: Oxford University Press.

———. (2006). *The Innate Mind, Volume 2: Culture and Cognition*. Oxford: Oxford University Press.

———. (2007). *The Innate Mind, Volume 3: Foundations and the Future*. Oxford: Oxford University Press.

Cherniak, C. (2005). Innateness and Brain-Wiring Optimization: Non-Genomic Nativism, In A. Zilhao (ed.), *Evolution, Rationality and Cognition*. London: Routledge.

Chomsky, N. (1955/1975). *The Logical Structure of Linguistic Theory*. Chicago: Chicago University Press.

———. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

———. (1966). *Cartesian Linguistics*. New York: Harper and Row.

———. (1968/2006). *Language and Mind*. Berkeley: University of California Press.

———. (1971). *Problems of Knowledge and Freedom: The Russell Lectures*. New York: Pantheon.

———. (2005). Three factors in language design. *Linguistic Inquiry* 36: 1–22.

———. (2007). Biolinguistic explorations: Design, development, evolution. *International Journal of Philosophical Studies* 15: 1–21.

Churchland, P. (1986), *Neurophilosophy*. Cambridge, MA: MIT Press.

Collins, J. (2005). Nativism: In defense of a biological understanding. *Philosophical Psychology* 18: 157–77.

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

———. (2009). Why Isn't Stich an Eliminativist?. In D. Murphy and M. Bishop (eds.), *Stephen Stich and His Critics*. Oxford: Blackwell.

Cracraft, J. (2000). Species Concepts in Theoretical and Applied Biology: A Systematic Debate with Consequences. In Q. Wheeler and R. Meier (eds.), *Species Concepts and Phylogenetic Theory*. New York: Columbia University Press.

Crane, T. (1992). The Non-Conceptual Content of Experience. In T. Crane (ed.), *The Contents of Experience*. Cambridge: Cambridge University Press.

Cudworth, R. (1678/1964). *The True Intellectual System of the Universe*. London, 1678. Facsimile reprint: Stuttgart-Bad Canstatt: Friedrich Frommann Verlag.

Damasio, A. (1994). *Descartes' Myth*: *Emotion, Reason and the Human Brain.* New York: Putnam.

Descartes, R. (1641a/1911). Meditations. In E. Haldane and G. Ross (eds. and trans.), *Philosophical Works of Descartes*, vol. 1. Cambridge: Cambridge University Press.

———. (1641b/1911). Author's Reply to Fifth Set of Objections. In E. Haldane and G. Ross (eds. and trans.), *Philosophical Works of Descartes*, vol. 2. Cambridge: Cambridge University Press.

———. (1647/1911). *Notes Against a Program*. In E. Haldane and G. Ross, G. (eds. and trans.), *Philosophical Works of Descartes*, vol 1. Cambridge: Cambridge University Press.

Devitt, M. (2011). Experimental semantics. *Philosophy and Phenomenological Research* 82: 418–35.

Devitt, M., and Sterelny, K. (1987/1999), *Language and Reality: An Introduction to the Philosophy of Language*, rev. 2nd ed. Cambridge, MA: MIT Press.

Dretske, F. (1980), *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

———. (1987). *Explaining Behavior*. Cambridge, MA: MIT Press.

Earman, J., Glymour, C., and Mitchell, S. (eds.) (2003). *Ceteris Paribus Laws*. Berlin: Springer.

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

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

———. (1981). The present status of the innateness controversy. In J. Fodor, *Representations: Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

———. (1991). *A Theory of Content*. Cambridge, MA: MIT Press.

———. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

———.(2001). Doing without what's within: Fiona Cowie's critique of nativism. *Mind* 100: 99–148.

———. (2008). *LOT2: The Language of Thought Revisited*. Oxford: Oxford University Press.

Fodor, J., and Lepore, E. (1992). *Holism: A Shopper's Guide*. Oxford: Blackwell.

Frege, G. (1884/1950). *The Foundations of Arithmetic*, translated by J. L. Austin. Oxford: Blackwell.

———. (1892/1966). On Sense and Reference. In P. Geach and M. Black (eds.), *Translations from the Works of Gottlob Frege.* Oxford: Blackwell.

Gelman, S. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*. New York: Oxford University Press.

Godfrey-Smith, P. (2007). Innateness and genetic information. In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind, Volume 3: Foundations and the Future*. Oxford: Oxford University Press.

Goldin-Meadow, S. (2003). *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us about How All Children Learn Language*. New York: Psychology Press.

Goodman, N. (1949). On likeness of meaning. *Analysis* 10: 1–7.

———. (1954/1983). *Fact, Fiction and Forecast*, 3rd ed. Cambridge, MA: Harvard University Press.

Gould, J., and Marler, P. (1991). Learning by Instinct. In D. Mock (ed.), *Behavior and Evolution of Birds*. San Francisco: Freeman.

Greenberg, M., and Harman, G. (2006). Conceptual Role Semantics. In B. Smith and E. Lepore (eds.), *Oxford Handbook of Philosophy of Language*. Oxford: Oxford University Press.

Griffiths, P. (1997). *What Emotions Really Are*. Chicago: University of Chicago Press.

———. (2002). What is innateness? *The Monist* 85: 70–85.
———. (2009). The distinction between innate and acquired characteristics. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/innate-acquired/.
Griffiths, P., and Gray, R. (1994). Developmental systems and evolutionary explanations. *Journal of Philosophy* 91: 277–304.
Griffiths, P., Machery, E., and Lindquist, S. (2009). The vernacular concept of innateness. *Mind & Language* 24: 605–30.
Hart, W. D. (1975). Innate Ideas and A Priori Knowledge. In S. Stich (ed.), *Innate Ideas*. Berkeley: University of California Press.
Hempel, C. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.
Herbert of Cherbury (1624/1937). *De Veritate*, translated by M. H. Carré. Bristol, UK: University of Bristol Press.
Horwich, P. (1998). *Meaning*. Oxford: Oxford University Press.
———. (2005). *Reflections on Meaning*. Oxford: Oxford University Press.
Hume, D. (1734/1978), *A Treatise of Human Nature*, 2nd ed., edited by L. A. Selby-Bigge and P. H. Nidditch. Oxford: Oxford University Press.
Jackendoff, R. (1983), *Semantics and Cognition*. Cambridge, MA: MIT Press.
———. (1992), *Languages of the Mind: Essays on Mental Representation*. Cambridge, MA: MIT Press.
Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
James, W. (1890/1983). *The Principles of Psychology*. Cambridge, MA: Harvard University Press.
Jerne, N. (1985). The generative grammar of the immune system. *Science* 229: 1057–59.
Johnson, M. (1997). *Developmental Cognitive Neuroscience*. Oxford: Blackwell.
Kant, I., (1787/1934). *Critique of Pure Reason*, 2nd ed., Translated by Norman Kemp Smith. New York: St. Martin's Press.
Katz, J. (1990). *The Metaphysics of Meaning*. Oxford: Oxford University Press.
Keil, F. (1989). *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.
Kitcher, P. (1996). *The Lives to Come*. New York: Simon & Schuster.
Kripke, S. (1972/80), *Naming and Necessity.* Cambridge, MA: Harvard University Press.
Laurence, S., and Margolis, E. (2002), Radical concept nativism. *Cognition* 86: 25–55.
Lehrman, D. (1953). A critique of Konrad Lorenz's theory of instinctive behaviour. *The Quarterly Review of Biology* 28: 337–63.
———. (1970). Semantic and Conceptual Issues in the Nature-Nurture Problem. In L. Aronson, E. Tobach, D. Lehrman, and J. Rosenblatt (eds.), *Development and Evolution of Behavior*. San Francisco: Freeman.
Lewis, D. (1972/1980). Psychophysical and Theoretical identifications. In N. Block (ed.), *Readings in the Philosophy of Psychology*, vol. II. Cambridge, MA: Harvard University Press.
Lewontin, R. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics* 26: 400–411.
Locke, J. (1690/1975). *An Essay Concerning Human Understanding*. Oxford: Oxford University Press.
Loewer, B. (1996). A Guide to Naturalizing Seemantics. In C. Wright and B. Hale (eds.), *A Companion to Philosophy of Language*. Oxford: Blackwell.
Lorenz, K. (1957). The Nature of Instincts. In C. H. Schiller (ed.), *Instinctive Behavior*. New York: International University Press.

Machery, E., Mallon, R., Nichols, S, and Stich, S. (2004). Semantics, cross-cultural style. *Cognition* 92: 1–12.
Macnamara, J. (1986). *Border Dispute: The Place of Logic in Psychology*. Cambridge, MA: MIT Press.
Mallon, R., and Weinberg, J. (2006). Innateness as closed process invariance. *Philosophy of Science* 73: 323–44.
Mameli, M. (2004). Nongenetic selection and nongenetic inheritance. *British Journal for the Philosophy of Science* 55: 35–71.
Mameli, M., and Bateson, P. (2006). Innateness and the sciences. *Biology and Philosophy* 21: 155–88.
Margolis, E. (1998/1999). How to acquire a concept. Reprinted in E. Margolis and S. Laurence (eds.), *Concepts: Core Readings*. Cambridge, MA: MIT Press.
Margolis, E. and Laurence, S. (eds.) (1999). *Concepts: Core Readings*. Cambridge, MA: MIT Press.
———. (2007). The ontology of concepts—abstract objects or mental representations? *Noûs* 41: 561–93.
Medin, D., and Atran, S. (eds.). (1999). *Folkbiology*. Cambridge, MA: MIT Press.
Mehler, J., and Dupoux, E. (1994). *What Infants Know: The New Cognitive Science of Early Development*, translated by P. Southgate. Oxford: Blackwell.
Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81–97.
Miller, G. and Johnson-Laird, P. (1976). *Language and Perception.* Cambridge, MA: Harvard University Press.
Millikan, R. (1984). *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.
AQ: 1 Moravcsik, J. (1975)
Murphy, G., and Medin, D. (1985/1999). The role of theories in conceptual coherence. In E. Margolis and S. Laurence (eds.), *Concepts: Core Readings*. Cambridge, MA: MIT Press.
Oyama, S. (2000). *Evolution's Eye: A Systems View of the Biology-Culture Divide*. Durham, NC: Duke University Press.
Pagin, P. (2006). Meaning holism. In B. Smith and E. Lepore (eds.), *Oxford Handbook of Philosophy of Language.* Oxford: Oxford University Press.
Peacocke, C. (1992). *Concepts.* Cambridge, MA: MIT Press.
Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic Books.
———. (1980). Language within Cognition: Schemes of Action and Language Learning. In M. Piatelli-Palmerini (ed.), *Language and Learning*. Cambridge: Harvard University Press, 163–83.
Peirce, C. (1903/1998). A syllabus of certain topics of logic. In *Essential Peirce*, vol. 2. Bloomington: Indiana University Press.
Pietroski, P., and Rey, G. (1995). When other things aren't equal: Saving *ceteris paribus* laws from vacuity. *British Journal for the Philosophy of Science* 46: 81–110.
Pinker, S. (2007). *The Stuff of Thought*. New York: Viking.
Plato (1975). *Meno*. Relevant passages reproduced in S. Stich, *Innate Ideas*. Berkeley: University of California Press .
Prinz, J. (2002). *Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.
Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.
Putnam, H. (1962a/1975). It ain't necessarily so. *Journal of Philosophy* 54. Reprinted in his *Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press.

———. (1962b/1975). Dreaming and depth grammar. Reprinted in his *Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press.

Quine, W. (1953). Two Dogmas of Empiricism. In W. Quine, *From a Logical Point of View and Other Essays*. Cambridge, MA: Harvard University Press.

———. (1956/1976). Carnap and Logical Truth. Reprinted in W. Quine, *Ways of Paradox and Other Essays* (revised ed.).

———. (1960). *Word and Object*. Cambridge, MA: MIT Press.

———. (1969). Natural Kinds. In W. Quinte, *Epistemology Naturalized*. New York: Columbia University Press.

Ramsey, F. (1929/1990). Theories. In F. Ramsey, *Philosophical Papers*, edited by D. Mellor. Cambridge: Cambridge University Press.

Rey, G. (1983/1999). Concepts and stereotypes. In E. Margolis and S. Laurence (eds.), *Concepts: Core Readings*. Cambridge, MA: MIT Press.

———. (1985). Concepts and conceptions. *Cognition* 19: 297–303.

———. (2005). Philosophical Analyses as Cognitive Psychology. In H. Cohen and C. Lefebvre (eds.), *Handbook of Categorization in Cognitive Science.* Amsterdam: Elsevier.

———. (2008). The analytic/synthetic distinction. *Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/entries/analytic-synthetic/.

———. (2009). Concepts, Defaults, and Internal Asymmetric Dependencies: Distillations of Fodor and Horwich. In N. Kompa, C. Nimtz, and C. Suhm (eds.), *The A Priori and Its Role in Philosophy.* Paderborn: Mentis, 185–204.

Rosch, E. (1973). On the Internal Structure of Perceptual and Semantic Categories. In T. E. Moore (ed.), *Cognitive Development and Acquisition of Language*. Amsterdam: Academic Press.

Samuels, R. (1998). What brains won't tell us about the mind: A critique of the neurobiological argument against representational nativism. *Mind & Language* 13: 548–70.

———. (2002). Nativism in cognitive science. *Mind & Language* 17: 233–65.

———. (2004). Innateness in cognitive science. *Trends in Cognitive Sciences* 8: 136–41.

———. (2007). Is Innateness a Confused Concept? In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind, Volume 3: Foundations and the Future*. Oxford: Oxford University Press.

Scholz, B., and Pullum, G. (2006). Irrational Nativist Exuberance. In R. Stainton (ed.), *Contemporary Debates in Cognitive Science*. Oxford: Blackwell.

Segal, G. (2007). Poverty of Stimulus Arguments Concerning Language and Folk Psychology. In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind, Volume 3: Foundations and the Future*. Oxford: Oxford University Press.

Siegal, M., and Bergman, A. (2002). Waddington's canalization revisited: Developmental stability and evolution. *PNAS* 9: 10528–32.

Smith, E., and Medin, D. (1981/1999). The Exemplar View (ch. 3 of their *Categories and Concepts*). In E. Margolis and S. Laurence (eds.), *Concepts: Core Readings*. Cambridge, MA: MIT Press.

Sober, E. (1988). Apportioning causal responsibility. *Journal of Philosophy* 85: 303–18.

———. (1998). Innate Knowledge. In W. Craig (ed.), *Routledge Encyclopedia of Philosophy, Volume 4*. London: Routledge.

———. (2001). Separating Nature and Nurture. In D. Wasserman (ed.), *Genetics and Criminal Behavior*. Cambridge: Cambridge University Press.

Stalker, D. (1994). *Grue: The New Riddle of Induction*, Chicago: Open Court.

Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.
Sterelny, K. (1989). Fodor's nativism. *Philosophical Studies*, 55: 119–141.
———. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Oxford: Blackwell.
Sterelny, K., and Griffiths, P. (1999). *Sex and Death*. Chicago: University of Chicago Press.
Stich, S. (ed.) (1975a). *Innate Ideas*. Berkeley: University of California Press.
———. (1975b). The Idea of Innateness. In S. Stich (ed.), *Innate Ideas*. Berkeley: University of California Press.
Thompson, D. (1917). *On Growth and Form*. Cambridge: Cambridge University Press.
Tooby, J., and Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology* 11: 375–424.
Tooby, J., Cosmides, L., and Barrett, C. (2005). Resolving the Debate on Innate Ideas: Learnability Constraints and the Evolved Interpretation of Motivational and Conceptual Functions. In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind, Volume 1: Structure and Contents*. Oxford: Oxford University Press.
Waddington, C. H. (1957). *The Strategy of the Gene*. London: Routledge.
———. (1975). *The Evolution of an Evolutionist*. Ithaca, NY: Cornell University Press.
Waters, K. (2007). Causes that make a difference. *Journal of Philosophy*, 104: 551–79.
Wendler, D. (1996). Innateness as an explanatory concept. *Biology and Philosophy* 11: 89–116.
Wilkin, A. (2003). Canalization and Genetic Assimilation. In B. Hall and W. Olson (eds.), *Keywords and Concepts in Evolutionary Developmental Biology*. Cambridge, MA: Harvard University Press.
Wittgenstein, L. (1953). *Philosophical Investigations*. New York: Macmillan.
Yang, C. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences* 8: 451–56.
Zalta, E. (2001). Fregean senses, modes of presentation, and concepts. *Philosophical Perspectives* 15: 333–59.

Author Query:

AQ 1: Please provide a corresponding entry in the reference list for Moravcsik (1975)