Mind & Language, ISSN: 0268-1064
Vol. 13. No. 3 September 1998, pp 421–449.

Theory Theory to the Max*

STEPHEN STICH AND SHAUN NICHOLS

1. Introduction: The Theory that Generates Shocked Incredulity

One approach to interdisciplinary theorizing is cautious and deferential, ever aware of the dangers involved in making pronouncements about matters on which one is hardly an expert. Typically, those who adopt this strategy propose tentative and elaborately nuanced theories whose complexity echoes the complexity of the phenomena to be explained. At the other end of the spectrum are the brash and provocative interdisciplinary theorists who do not hesitate to defend ambitious, uncomplicated theories that have far reaching implications for fields that are not their own. We doubt there is any way of knowing in advance which of these approaches will be more successful. But, for us at least, the provocative approach is usually a lot more fun. Nor is that its only virtue. For when a theory is bold and straightforward it is usually much easier to see the problems it confronts, and to learn from them.

Those who share our taste for audacious, uncluttered, far-reaching theories will find no shortage of instructive provocation in Gopnik and Meltzoff's book and in a pair of closely related articles that have recently appeared in *Philosophy of Science*.[1] 'The theory theory' is the label that G and M adopt for their view, and 'the central idea of this theory is that the processes of cognitive development in children are similar to, indeed perhaps even identical with, the processes of cognitive development in scientists' (GM, 3). As Gopnik sees it, 'the moral of [the] story is not that children are little scientists but that scientists are big children. Scientists and children both employ the same particularly powerful and flexible set of cognitive

*Critical Notice of Alison Gopnik and Andrew N. Meltzoff, *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press, 1997. Pp. xvi + 268.

We are grateful to Stephen Downes, Peter Godfrey-Smith and Jerry Fodor for a number of helpful suggestions.

Address for correspondence: Stephen Stich, Department of Philosophy, Davison Hall/Douglass Campus, Rutgers University, 26 Nichol Avenue, New Brunswick, NJ 08901–2882, USA.

Email: stephen-stich@email.msn.com, nichols@cofc.edu.

[1] Gopnik (1996a and 1996b). We will adopt the following abbreviations: Ga = Gopnik, 1996a; Gb = Gopnik, 1996b; GM followed by a page number = Gopnik and Meltzoff, 1997; G and M without page numbers = Gopnik and Meltzoff (the authors, not the book).

devices' (Ga, p. 486). '[E]veryday cognition, on this view, is simply the theory that most of us most of the time have arrived at when we get too old and stupid to do more theorizing. . . . We might think of our enterprise as scientists as the further revision of the theory by the fortunate, or possibly just childish, few who are given leisure to collect evidence and think about it' (GM, p. 214). Indeed, Gopnik goes on to suggest that 'the greatness of individual scientists would literally come from their childishness' (Gb, p. 561).

Suggestions like these have been broached before in the literature by these authors and others (Gopnik, 1984 and 1988; Gopnik and Wellman, 1992 and 1994; Wellman, 1985 and 1990; Wellman and Gelman, 1992), though, as Gopnik reports, the ideas are often greeted with 'shocked incredulity' by 'scientists, philosophers and psychologists, particularly those with limited experience of anyone younger than a freshman' (Ga, p. 486). In *Words, Thoughts and Theories*, G and M attempt to undermine this scepticism by articulating their position in detail and showing 'how it can generate specific predictions, predictions that are not made by other theories' (GM, p. 4). They also survey an impressive range of data about cognitive development in infancy and childhood, and argue that the theory theory can explain these data. Their book is, by far, the most detailed defence that has yet been offered for the theory theory. Moreover the position they advocate is, in a number of ways, an unabashedly *extreme* version of the theory theory. It's Theory Theory to the Max!

To get a bit clearer on what G and M's version of the theory theory claims, and to see why it is so bracingly simple and radical, it will be useful to sketch several of the dimensions on which their theory is more radical than other views to which the 'theory theory' label has been applied. That will be our project in the section to follow. In section 3, we will briefly consider some of the evolutionary considerations that G and M offer in support of their theory, and note some further evolutionary considerations that pull in the opposite direction. In section 4, by far our longest section, we'll provide a more detailed statement of G and M's theory, sketch some of the facts which, they claim, the theory can explain, and take a critical look at their arguments aimed at showing that their theory is better than the competition. In the final section, we'll explain why their theory requires them to adopt some extremely controversial views about the nature of scientific change.

2. *Far Out: Some Dimensions Along Which Theory Theories Differ*

To the best of our knowledge, the term 'theory theory' was first used by Adam Morton (1980) to characterize a cluster of views, in philosophy and psychology, about the ways in which normal adults go about the business of predicting and explaining the actions of other people and attributing mental states to them. In all of these activities, according to the sort of theory theory that Morton had in mind, people are relying on 'a fairly extensive commonplace psychological theory, concerned both with dispositional traits

such as those of character and mood and with the intentions that produce action.' (Morton; 1980, p. 13) Not everyone thinks that our 'folk psychological' capacity to predict and explain behaviour and to attribute mental states relies on a commonsense theory. Behaviourists don't, of course. Neither do advocates of the 'simulation theory,' who maintain that in all these activities we gain information about other people by running some of our own psychological mechanisms 'off-line'.[2] But even among those who think that some sort of commonsense theory is a crucial part of the explanation of our everyday folk psychological skills, there is lots of room for disagreement on what, exactly, counts as a theory. And this is the first dimension on which G and M stake out an extreme position. For Morton, 'not any collection of beliefs forms a theory' (p. 6); a certain degree of unity is required, and the terms of a theory must be semantically interdependent. But if there is a commonsense psychological theory underlying adult folk psychological abilities, it needn't, Morton insists, look much like a *scientific* theory. 'There are', he reminds us, 'theories outside science too' (p. 6). Others, including one of the authors of this paper, have taken an even more permissive view. Our use of commonsense psychological terms, Stich once claimed, 'is governed by a loose knit network of principles, platitudes and paradigms which constitute a sort of folk theory' (Stich, 1983, p. 1). And in more recent work, where the focus was the plausibility of off-line simulation theories, we have used the term 'theory theory' in a way that (contra Morton) would count just about any collection of beliefs as a theory (Stich and Nichols, 1992 and 1995). But G and M have no truck with this insipid inclusiveness. They offer a detailed account of the structural, functional and dynamic features which, they claim, characterize most scientific theories (GM, pp. 32–41), and they go on to claim that the cognitive structures subserving folk psychological skills in both children and adults have *all* of these features. More generally, when G and M claim that a given body of knowledge in any domain is theoretical, they mean that it has all of the characteristics they attribute to scientific theories.[3]

A second way in which G and M embrace a stronger position than many others who would call themselves 'theory theorists' turns on the range of commonsense capacities which are said to be subserved by a theory. For Morton, as we've seen, the theory theory is a hypothesis about commonsense psychological capacities. But G and M extend the hypothesis to cover three

---

[2] For more on the simulation theory, see the essays in Davies and Stone, 1995a and 1995b, and in Carruthers and Smith, 1996.

[3] We should note that there is no substantive disagreement between G and M and ourselves here. In our defence of the theory theory in Stich and Nichols, 1992 and 1995, we were concerned to argue that adult folk psychological skills are subserved by a mentally represented knowledge structure—a body of folk psychological information—rather than by a process of off-line simulation. And, of course, G and M agree. Given our polemical goals, there was no need for us to take a stand on how similar this mentally represented knowledge structure is to a scientific theory, and thus we simply left the matter open.

large domains of commonsense knowledge. The first of these, which they call the 'theory of object appearances', includes knowledge about the movement of objects, the properties of objects, the spatial relations between stationary objects and the perceptual relations between observers and objects (GM, p. 78). On G and M's account, 'the theory of object appearances is an interesting bridge between 'folk physics' and 'folk psychology'. It involves ideas about objects and the relations among them, but, just as crucially, it involves ideas about the way that we and others perceive objects' (GM, p. 78). The second domain in which G and M maintain that commonsense knowledge is theoretical is what they call 'the theory of action'. This includes much of what is usually classified as commonsense psychological information, including information about how beliefs and desires arise and change and how they interact to produce actions. It also includes information about the ways in which actions have effects both on the physical world and on people. The third domain that G and M discuss is 'the theory of kinds', which specifies the principles according to which objects are to be grouped together into kinds and the sorts of inferences to be drawn from the fact that objects are of the same kind. Though these are the three domains that G and M treat in detail, they do not preclude the possibility that knowledge in other domains is also theoretical[4]. On the other hand, they make it clear that they do not think all commonsense knowledge is theoretical. Rather, they expect that scripts, narratives, empirical generalizations and other forms of information packaging will all have a role to play in explaining some aspects of commonsense knowledge.

Thus far we've been considering what G and M's version of the theory theory has to say about the knowledge structures that underlie a range of normal adult capacities. But since both Gopnik and Meltzoff are developmental psychologists it is not surprising that the boldest and most radical aspect of their theory is the position they take on developmental questions—questions about how these adult knowledge structures emerge. Here there are two issues to consider. First, when exactly does our knowledge in the relevant domains *start* being theoretical? As we trace a person's knowledge in these domains back from adulthood to adolescence through childhood to infancy, when do the first theories appear? The astonishing answer that G and M propose (with obvious delight) is that the child's knowledge in the three domains they consider is *always* theoretical; right from the start it has all the features that are fundamental to scientific theories.

> At this point some of the psychologists, scientists and philosophers who were crying out . . . may well be doing so again; 'Surely, you cannot think it is theories all the way down!' Well, yes, actually I do think it is theories all the way down . . . Infants seem to have

[4] The domains they mention include 'folk economics' (GM, p. 174), 'hunter-gatherer folk botany' and 'Aborigine geography' (Gb, p. 561).

> innate knowledge . . . and this knowledge is theory-like . . . (When we say this knowledge is innate we do not mean this in the philosophical sense, which is that neither the philosopher in question nor any of the guys down the hall could think of a way to learn it. We mean that it has been demonstrated in 42-minute-old infants.) (Ga, p. 510)

Most previously published applications of the theory theory have involved older children, including some who are actively learning scientific concepts. But these studies have no direct bearing on what G and M dub 'initial state nativism'—the claim that infants are *born* knowing theories. One of the main aims of their book 'is to apply the theory theory to explain what we know about infancy and very early childhood' (GM, p. 4), and thus much of the evidence they recount deals with cognitive development in the first 36 months of life.

A second developmental question focuses on the mechanisms that are responsible for the theoretical revisions and replacements that (they maintain) occur during the course of cognitive development in childhood. What are these mechanisms? How exactly do they work? Here G and M concede that they do not have a complete story to tell. (GM, pp. 218 ff) What they do claim, as we've already noted, is that the mechanisms responsible for theory change in infants and children are exactly the same as the mechanisms that are responsible for theory change in science. According to G and M, human beings have only one set of mental mechanisms for theory revision. These mechanisms play a major role in the cognitive changes that mark everyone's infancy and childhood. In the privileged few who have the leisure to pursue science as adults, the mechanisms continue to be used.

3. *Some Evolutionary Puzzles*

One virtue that G and M claim for their version of the theory theory is that it solves 'an interesting evolutionary puzzle'. Everyone agrees that humans (at least *some* humans) have the capacity to do science. And few would challenge the claim that in doing science people use a flexible and powerful set of cognitive abilities. But, G and M ask, 'Where did the particularly powerful and flexible devices of science come from? After all, we have only been doing science in an organized way for the last 500 years or so; presumably they didn't evolve so that we could do that' (GM, p. 18 and Ga, p. 489). The answer they suggest is that

> many of these cognitive devices are involved in the staggering amount of learning that goes on in infancy and childhood. Indeed, we might tell the evolutionary story that these devices evolved to allow human children, in particular, to learn (GM, p. 18, and Ga, p. 489).

> From an evolutionary point of view, three of the most distinctive features of human beings are the plasticity of their behavior, their ability to adapt to an extremely wide variety of environments and their long, protected immaturity. Equipping human children with particularly powerful and flexible cognitive devices, devices that are good at constructing accurate representations of new and unexpected worlds, might be an important part of this evolutionary strategy. We might indeed think of childhood as a period when many of the requirements for survival are suspended, so that children can concentrate on acquiring a veridical picture of the particular physical and social world in which they find themselves. Once they know where they are, as it were, they can figure out what to do. On this view we might think of infancy as a sort of extended stay in a Center for Advanced Studies, with even better food delivery systems . . .
>
> . . . these powerful theory formation abilities continue to allow all of us at some times, and some of us, namely professional scientists, much of the time, to continue to discover more and more stuff about the world around us. On this view science is a kind of spandrel, an epiphenomenon of childhood. (Ga, p. 490)

This proposed solution to the evolutionary puzzle gives rise to two further puzzles, one of which Gopnik acknowledges and addresses with considerable ingenuity. The other goes unnoticed.

The puzzle that Gopnik acknowledges is what Giere (1996, p. 539) calls 'the 1492 problem'. 'Science as we know it'. Giere notes, 'did not exist in 1492'. But if G and M are right, then the cognitive devices that give rise to science have been part of our heritage since the Pleistocene. *So why have humans only been doing science for the last 500 years?* The answer, according to Gopnik, is mostly a matter of the availability of relevant evidence. 'My guess is that children, as well as ordinary adults, do not . . . systematically search for evidence that falsifies their hypotheses, though . . . they do revise their theories when a sufficient amount of falsifying evidence is presented to them. In a very evidentially rich situation, the sort of situation in which children find themselves, there is no point in such a search; falsifying evidence will batter you over the head soon enough' (Gb, p. 554). Now what happened about 500 years ago, Gopnik maintains, is that as a result of various historical and social factors a few thinkers found themselves confronted with unprecedented amounts of new evidence relevant to several venerable questions like: Why do the stars move as they do? New technology was one reason for the availability of new evidence; telescopes were invented. Other technological and social changes greatly facilitated communication allowing 'a mathematician in Italy to know what an astronomer has seen in Denmark' (Gb, p. 554). Greater leisure (at least for a few) was yet another factor, and so too was the emergence of the experimental method which motivated those who adopted it to systematically seek out new and potentially falsifying evidence. All of this, and perhaps other factors as well, created an environ-

ment in which the theory revision mechanisms that natural selection had designed to enable children to cope with 'new and unexpected worlds' might begin functioning actively in adulthood, long past the stage in life in which they would have made their principal contribution to fitness in the environment in which our ancestors evolved.

There is, we think, some plausibility to this story. It meshes nicely with recent accounts that stress the importance of environmental instability in hominid evolution. (See, for example, Potts, 1996.) If, as the evidence suggests, climactic changes became much more frequent during the last 2.8 million years, hominids that were good at learning to deal with radically changed environmental conditions might well have had a considerable selective advantage. But there is also a significant, and unacknowledged, tension between this account and one of the more interesting and extreme features of G and M's view. Recall that according to G and M 'it's theories all the way down' and that even the very early development of the child's knowledge about object appearances, actions and kinds is driven by a process of theory revision—the very same process whose evolutionary function is to enable older children (and scientists) to acquire 'a veridical picture of the particular physical and social world in which they find themselves' and thus 'to adapt to an extremely wide variety of environments' (Ga, p. 490). Now what is puzzling about this is that just about *all* the knowledge that children are acquiring in the early stages of this process—knowledge about the continuous movement of objects, for example, or about the spatial relations that must obtain between an agent and an object if the agent is able to see the object, or about the way in which an agent's desires lead to actions—deals with aspects of the environment which are *not in the least variable*. These facts have been fixed and unchanging since long before the emergence of hominids or primates. Indeed, some of these facts (particularly those about the movements and interactions of middle-sized physical objects) have presumably obtained in the environment of every organism that has ever existed on Earth. Moreover, it is also the case that knowledge of these facts would be adaptive even to creatures which did not have to cope with a highly variable environment and which (because of this, or for whatever reason) never evolved anything like the human capacity to develop veridical theories about a wide variety of environments. Thus it is hardly surprising that, as Carey and Spelke (1996) note, infants and primates reason similarly about objects, or that young children and other mammals construct similar representations of space. There is even some evidence suggesting that 2-day-old chicks perceive the complete shapes of partly hidden objects and track the location of fully hidden objects in much the same way that human infants do.

Obviously, the mechanism responsible for the young chicken's acquisition of the ability to track the location of hidden objects must be very different from the powerful and flexible cognitive devices that enable humans to construct accurate representations of new and unexpected worlds. Moreover, it is overwhelmingly likely that our own primate and pre-primate ancestors

had the knowledge necessary to deal with the movements and trajectories of physical objects long before the appearance of the sort of theory revision mechanisms posited by G and M. So there must have been some mechanisms in place in our primate forebears which explained their possession of this knowledge, and those mechanisms could not have been the flexible and powerful theory revision devices. Yet if G and M are right, the theory revision devices *are* the mechanisms responsible for the acquisition of much the same knowledge in human infants and young children. But it is puzzling, to put it mildly, how this change might have come about. Why would natural selection abandon a perfectly good and extremely reliable system for generating knowledge about certain utterly stable and unchanging parts of the environment, and assign the task to a new, more fallible system whose main adaptive virtue is that it can gain knowledge about highly variable parts of the environment? If the older, inflexible system ain't broke, why on earth would natural selection fix it? Wouldn't it make more sense for natural selection to retain the older system for the early acquisition of knowledge about stable features of the world, and arrange for the theory revision device to kick in later on in development when children must figure out which of the many possible environments they happen to have ended up in? A two-tiered system of this sort could still tell the story G and M tell about science being a spandrel, though it would be an epiphenomenon of a somewhat later stage of childhood. What this two-tiered hypothesis rejects is G and M's radical contention that it's theories (and theory revision) all the way down.

These evolutionary considerations hardly constitute a knockdown argument against G and M's version of the theory theory. There are lots of stories to be told about why natural selection might dump the system that yields knowledge about spatial relations and the motions of physical objects in non-human primates and replace it with a theory revision system of the sort that G and M propose. Perhaps it's easier to integrate primitive and more sophisticated knowledge if the two are generated by the same system; perhaps maintenance costs are lower if there is only one system, rather than two.[5] This game is easy to play. The conclusion we would draw is that while G and M's theory is compatible with lots of the facts that need explaining, the fit is far from perfect, and there are alternative theories that do at least as good a job at explaining the facts. This conclusion will become a leitmotif of the section to follow, where our central theme will be that G and M do

[5] Indeed, there is reason to think that natural selection often proceeds in just the opposite direction. The so-called 'Baldwin effect' is a process in which Darwinian selection can mimic aspects of the Lamarckian inheritance of acquired characteristics. In Baldwinian evolution, learned traits which are reliably adaptive in a given environment are gradually genetically assimilated because mutations that result in some or all of the previously learned information becoming innate will, under appropriate circumstances, be favoured by natural selection. For useful discussions of the Baldwin effect, see Godfrey-Smith, 1991, section 6.5, and Deacon, 1997, ch. 11. For an elegant and influential attempt to model the Baldwin effect, see Hinton and Nolan, 1987).

not make a convincing case for the claim that their theory does a better job than the competition at explaining the evidence.

4. The Evidence, The Theory and The Competition

The central claim of G and M's theory is that the processes of cognitive development in children, including very young children, are similar or identical to the processes of cognitive development in science. To make a case for this claim, G and M adopt the following strategy. First, they give an account of the processes of cognitive development in science. Next, they recount a great deal of experimental and anecdotal evidence about cognitive development in children, and argue that all of these facts are compatible with their theory—if the theory were true it would explain the experimental and anecdotal findings. But, as G and M are aware, this by itself would only be enough to show that their theory is a serious contender. In order to argue that their theory offers *the best* explanation of the available facts they must, and do, argue that there are facts which competing theories cannot comfortably explain. It's our contention that their arguments for this last claim are singularly unpersuasive. To explain why we find their arguments unconvincing, we will proceed as follows. First, we'll sketch the account of scientific theories and their development that G and M adopt. Next, we'll offer a few illustrations of the sorts of evidence about young children that G and M claim can be explained by their theory. We'll then consider in some detail the arguments that G and M offer for the claim that their theory does a better job at explaining the facts than competing theories.

4.1. An Account of Scientific Theories and Their Development

There is, as G and M note, considerable disagreement among philosophers, historians and sociologists of science about just what scientific theories are and how they change. Confronted with this controversy, G and M propose to take 'the modest and emollient route of focusing on those features of theories that are most generally accepted across many different conceptions of science', since they want to be 'as mainstream and middle-of-the-road as possible' (GM, p. 33). This is, we think a fair characterization of much of what they say about science, though in section 5 we'll note that some of the claims their theory requires them to make about science are far from mainstream, and far from plausible.

In setting out their account of theories, G and M focus on three sorts of features—structural, functional and dynamic—which they claim are distinctive of scientific theories. The four 'static structural' features they mention are:

(1) Abstractness: 'theoretical constructs are typically phrased in a vocabu-

lary that is different from the vocabulary of the evidence that supports the theory' (GM, p. 34).

(2) Coherence: 'The entities postulated by a theory are closely, "lawfully," interrelated with one another' (GM, p. 35).

(3) Causality: 'in theories we appeal to some underlying causal structure that we think is responsible for the superficial regularities in the data. . . . The intratheoretic relations, the laws, are typically interpreted in a causal way . . . [and] the theoretical entities are seen to be causally responsible for the evidence (GM, p. 35).

(4) Ontological commitment: 'theories make ontological commitments and support counterfactuals . . .. [Thus] a test of theoreticity . . . is the nature of our surprise at violations of the theory. If we are committed to the theory, such violations strike us not only as surprising but as being impossible and unbelievable in an important and strong way. This differentiates theories from other types of knowledge' (GM, pp. 35–6).

The functional features, the things that theories do (or that people do with them) are:

(5) Prediction: 'A theory, in contrast to a mere empirical generalization, makes predictions about a wide variety of evidence, including evidence that played no role in the theory's construction' (GM, p. 36).

(6) Interpretation: 'theories strongly influence which pieces of evidence we consider salient or important' (GM, p. 37).

(7) Explanation: 'The coherence and abstractness of theories and their causal attributions and ontological commitments together give them an explanatory force lacking in mere topologies of, or generalizations about, the data' (GM, p. 38). And why do people seek explanatory theories? Evolution built us with the motivation to explain because that motivation leads us to build theories, and good theories enable us to deal more effectively with our environment. 'From an evolutionary point of view we might suggest that explanation is to cognition as orgasm (or at least male orgasm) is to reproduction' (GM, p. 38).

Though all these features play a role at one point or another in G and M's argument for the theory theory, much of the weight of their argument rests on the dynamic features of theories:

(8) Defeasibility: 'the most important thing about theories is what philosophers call their defeasibility. Theories may turn out to be inconsistent with the evidence, and because of this theories change' (GM, p. 39).

(9) The crucial role of counterevidence: 'Theories change as a result of a number of different epistemological processes. One particularly critical factor is the accumulation of counterevidence to the theory' (GM, p. 39).

(10) Characteristic intermediate processes: 'There are characteristic intermediate processes involved in the transition from one theory to another' (GM, p. 39).
   (a) Denial: 'The initial reaction of a theory to counterevidence may be a kind of denial. The interpretive mechanism of the theory may treat the counterevidence as noise, mess, not worth attending to' (GM, p. 39).
   (b) Auxiliary hypotheses: 'At a slightly later stage the theory may develop ad hoc auxiliary hypotheses designed to account superficially for the counterevidence. . . . But such auxiliary hypotheses often appear, over time, to undermine the theory's coherence. . . . The theory gets ugly and messy instead of being beautiful and simple. The preference for beautiful theories over ugly ones (usually phrased, less poetically, in terms of simplicity criteria) plays an additional major role in theory change' (GM, p. 39).
   (c) Appearance of an alternative: 'The next step requires an alternative model to the original theory. . . . The ability to fix on this alternative is the mysterious logic of discovery' (GM, p. 40).
   (d) Intense experimentation and observation: 'A final important dynamic feature of theory formation is a period of intense experimentation and/or observation. . . . The role that experimentation and observation play in theory change is still mysterious, but that it plays a role seems plain' (GM, p. 40).

4.2. The Theory Meets the Evidence

The version of the theory theory that G and M want to defend claims that children, including very young children, have and use theories with the structural and functional features sketched in (1)–(7), and that, in the course of development, these theories are replaced with better theories by a process that fits the pattern elaborated in (8)–(10). This is not the place to attempt a detailed summary of experimental findings that G and M recount, though much of this evidence is fascinating and surprising, and the tour they offer is well worth the price of the book even if you are no more convinced by their theory than we are. But while there can be little doubt that the results they discuss are intriguing and important, it is often a bit of a stretch to see why they think the evidence supports their theory.

Consider, for example, Meltzoff and Moore's (1983, 1989) amazing finding that infants as young as 42 minutes old are able to imitate facial gestures such as mouth opening, tongue protrusion and lip protrusion. Since infants cannot see their own faces, G and M argue that this capacity requires an innate, cross-modal mapping. 'There is', they maintain, 'an abstract representation, a kind of body scheme, that allows an innate mapping from certain

kinds of behavioral observations of others to certain kinds of perceptions of our own internal states. In particular, we innately map the visually perceived motions of others onto our own kinesthetic sensations' (GM, p. 129). All of this is plausible enough. But what bearing does it have on the claim that 'infants have innate knowledge . . . and this knowledge is theory-like' (Ga, p. 510). Well perhaps, as Gopnik suggests, the abstractness of the representation that subserves the cross-modal mapping shows that the child's innate knowledge has the first of the ten features on the list set out in 4.1. Gopnik also maintains that we can view these infants as 'drawing at least a primitive kind of inference and prediction' (Ga, p. 510) on the basis of this knowledge, thus getting us feature (5). We're inclined to think that this requires a rather generous interpretation of the facts. For what exactly does Gopnik think the infant is *predicting*? The most obvious suggestion is that he is predicting that kinesthetic sensations of a certain sort will produce behaviour similar to the behaviour he has just observed. And we find it a bit hard to take this proposal seriously. But even if we concede the point on feature (5), what about the other eight features? What reason do we have to suppose that the infant takes the entities posited by his theory (whatever exactly these are) to be 'closely, lawfully, related with one another', or that the infant is interpreting anything in a causal way, or that his knowledge supports counterfactuals, or that what he knows has an explanatory force? The answer, as best we can see, is that we have no reason at all to suppose any of this. Of course, this hardly shows that G and M's theory is mistaken. Perhaps the infant is making lots of causal interpretations, contemplating counterfactuals and generating orgiastic explanations one after another, and we just haven't yet figured out how to demonstrate this experimentally. Or perhaps, though the infant has a theory which *would* support all of these cognitive activities, he's just not much inclined to engage in any of them at this very tender age. The point here is that while G and M's theory may be broadly compatible with the evidence they cite, the theory claims vastly more than the evidence gives us any reason to believe. Moreover, the evidence is equally compatible with a wide range of alternative hypothesis which do not attribute to 42-minute-old infants a knowledge of theories with all the fundamental properties of scientific theories.

For a second example, let us consider G and M's intriguing proposal about the much studied A-not-B error that infants make between the ages of 9 and 12 months. During the first 9 months of life an infant's knowledge about the motions of objects evolves in complex ways. By 6 months gaze tracking experiments show that infants can project the visible trajectory of an object even when it disappears behind a screen. But they seem quite untroubled if an object smoothly moves from left to right, disappears behind the left edge of one screen, and reappears at the right edge of a second screen without ever appearing in the gap between the two screens. Also, at this age, they fail to search for an object, even a highly desirable one like a favourite toy, when it disappears behind a screen or a cloth. By 9 months the situation has changed considerably. While their gaze will still track the trajectory of an

object as it passes behind a screen, if the object fails to appear at the gap between one screen and another, their tracking will be disrupted and they will look back to the first screen. Also, at 9 months, they have little trouble finding an object hidden under a single cloth. If an object hidden under a cloth disappears, 'they even shake the cloth as if they expect the object to appear there. All this suggests that these infants genuinely postulate that the object will be behind the occluder where it disappeared' (GM, p. 95). Oddly, however, a 9-month-old will often make what is known as the *A-not-B error*: If the infant observes an object being hidden and recovered several times under cloth A and then observes the object being hidden under a different cloth, B, the infant searches persistently under cloth A rather than under cloth B.

The explanation that G and M propose for these and other facts is as follows. At 9 months, children have 'a rich and abstract conception of objects' (GM, p. 98). They 'believe that the object is at the place or trajectory at which it disappeared *and* that it is behind the occluder *and* that the fact that it is behind the occluder is what makes the object invisible' (GM, p. 95). However, there are many common situations which this theory simply does not handle. These include situations in which the object is 'invisibly displaced in the sense that it changes its trajectory and traces an invisible path of movement that cannot be extrapolated from its visible path' (GM, p. 97). This is what happens, for example, when a child throws her teddy bear out of her crib (and out of sight) and, with some help from Mom, it later ends up in the toy box. 'What can the baby do in these circumstances? There is a rule that can handle the teddy bear case and many others where invisibly displaced objects reappear at places that were not part of their original trajectory. Moreover, the rule doesn't require that you have a theory of invisible displacements. Many objects, especially in the child's world, have habitual locations. The rule is, "The object will be where it appeared before" ' (GM, p. 99). This rule 'will receive a great deal of empirical confirmation. We might think of it as an example of a purely empirical generalization. . . . But from the point of view of the adult theory, and arguably of any coherent theory, both this generalization and the theoretical prediction cannot simultaneously be true; the object can't be under both cloth A [where it appeared before] and under cloth B [where it disappeared]. . . . In fact, the evidence suggests that infants may also develop other often contradictory empirical generalizations in the 9–to–15-month period' (GM, pp. 99–100).

For G and M, the payoff in all of this is the link with (10a) on the list of the features of scientific theories: 'This situation is analogous to similar situations in science where, without a theory to resolve them, contradictory generalizations may proliferate. The 9-month-old is like a scientist who attempts to save the theory by adding ad hoc auxilliary hypotheses. The central theory of object movements leads to apparent anomalies when objects are invisibly displaced. The ad hoc rule "It will be where it appeared before" is invoked to deal with these anomalies' (GM, p. 100). By about 12 months, children stop making A-not-B errors, and this leads G and M to claim that they have

abandoned the ad hoc auxilliary hypothesis. 'Why do they abandon it? The fact that the rule is sometimes disconfirmed, as in the A-not-B situation, probably plays a part. We suspect, however, that the lack of consistency between this rule and the centrally developed theory of objects that leads to the rule "It will be where it disappeared" plays an even more important role. To return to our earlier account, the child is analogous to a scientist who is disturbed to discover that her ad hoc auxilliary hypotheses lead to contradictory predictions' (GM, p. 100).

As we see it, the appropriate conclusion to draw from this example is quite similar to the conclusion we drew from the previous one: the evidence that G and M cite is indeed broadly compatible with their strong version of the theory theory. However, here as before, one has to be rather generous in interpreting the theory to get it to square with the facts. One of these facts is that, as G and M note (GM, p. 96), under certain circumstances 9-month-olds will make A-not-B errors *even when the object at B is still clearly visible*. Presumably the explanation that they would offer for these cases is that the child's reliance on the ad hoc auxilliary hypothesis is so strong that it overrules not only theoretical predictions about invisible objects but also direct observations of visible ones. It's possible, we suppose, though surely it's a bit of a stretch. Moreover, as both G and M and Carey and Spelke (1996, pp. 521–2) note, there are other explanations in the literature that trace A-not-B errors to maturational changes in the brain structures subserving means/end planning and the inhibition of competing responses. Carey and Spelke also note that 'Diamond ... has shown that the developmental changes involving the A/not B errors of infants of 7 months and beyond, are mirrored, in parametric detail, by identical changes in 2- to 4-month-old rhesus monkey infants' (Carey and Spelke, 1996, p. 521). We doubt that G and M would explain this finding by attributing ad hoc hypotheses to rhesus monkeys, so they would have to argue that, despite the strong parallels, the processes underlying A-not-B errors in humans is radically different from the processes at work in monkeys.

We've considered only two examples drawn from the impressive body of evidence that G and M assemble. However, what we've said about these two cases can, we think, be said about much of the rest of their evidence: with a bit of squinting it can all be seen as broadly compatible with their theory. And the fact that their theory is compatible with such a large and varied body of developmental evidence is, we think, more than enough to establish it as a real contender. Despite the 'shocked incredulity' with which the theory is often greeted, G and M have made an impressive case that their strong, theories-all-the-way-down version of the theory theory must be taken very seriously indeed. But of course G and M would hardly be satisfied with this. They claim that their theory is not only a *serious* contender but the *leading* contender. To defend this much stronger claim, they have to argue that their theory is better than the available alternatives. And argue they do. As we see it, however, these arguments are by far the weakest parts of their case.

4.3. The Theory Meets the Competition

As G and M view it, the principal competitors to the theory theory fall into two broad categories. One of these, which they call 'empirical generalizations' includes 'scripts, narratives, connectionist nets, and other cognitive structures quite closely related to immediate experience' (GM, p. 50). Since they recognize that both the potential of connectionist models and the exact nature of the claims made for them are far from clear (GM, p. 218), G and M quite sensibly do not mount a detailed critique of such models. They do, however, argue that scripts, narratives and similar structures are simply insufficiently abstract to explain the wide range of predictions, including predictions about entirely novel cases, that young children can make about objects and actions. We think the case they make is quite convincing.

For the other category of competing theories, G and M use the label 'modules'. As they characterize them, modularity theories claim that 'representations of the world are not constructed from evidence in the course of development. Instead, representations are produced by innate structures, modules, or constraints that have been constructed in the course of evolution. These structures may need to be triggered, but once they are triggered, they create mandatory representations of input' (GM, p. 50). 'The classic examples of modules are the specialized representations and rules of the visual and syntactic systems' (GM, p. 51). Other examples that G and M discuss include the 'core knowledge' theory advocated by Spelke, Carey and others, Leslie's account of the 'theory of mind mechanism,' and Cosmides and Tooby's theory about the cognitive mechanism that subserves reasoning about permission and obligation. G and M do not claim that modules play no role in cognition. Quite the opposite. They advocate 'a kind of developmental pluralism: there are many quite different mechanisms underlying cognitive development' (GM, p. 49). However, they maintain that 'theory formation rather than these other mechanisms accounts for the particular cognitive and semantic phenomena' that they describe in their discussion of the theory of object appearances, the theory of action and the theory of kinds (GM, p. 50). They also make it clear that what divides them from modularity theorists is not a dispute over *nativism*, for 'while modules are innate, not all innate structures are modules' (GM, p. 51), and on their version of the theory theory, infants are born with innate *theories*. Unlike modules, these innate theories are 'defeasible; any part of them could be, and indeed will be, altered by new evidence' (GM, p. 51).

4.3.1. An Argument Based on the Static Properties of Theories How are we to tell whether the mental representations that underlie the child's developing skills in dealing with objects, agents and kinds are the product of innate modules or of processes of theory revision of the sort that G and M champion? There are, it seems, two sorts of evidence that we might consider. The first is evidence that bears on the synchronic or static features of these representations. Since G and M claim that it's theories all the way down, we

should expect that the representations subserving the child's skills with objects, agents and kinds always manifest the first seven features on the list set out in 4.1, and, as we have noted, G and M do cite some evidence that this is the case even very early on. However, when the goal is to show not merely that the theory theory is compatible with the evidence, but that it is better than the module alternative, evidence bearing on the static properties of the child's mental representations simply is not relevant. For while a modularity theorist *might* claim that the representations underlying one of these skills lacks some of the features (1)–(7), a modularity theorist might also claim that some innate modules exploit representations that are *identical with theories in all of their static properties*. So, while evidence bearing on the static features of the representations underlying skills with objects, agents and kinds could in principle falsify G and M's theory, such evidence could not falsify the modular alternative. Most of the time G and M see this point very clearly. For example, in discussing 'what kinds of evidence could differentiate between a modularity theory and a theory theory that includes innate theories', they note that

> it may be difficult, if not impossible, to distinguish these views by looking at a single static representational system. At least some of the structural and functional features of theories—their abstractness, coherence, and predictive interpretive force—can also be found in modules (GM, p. 52; for similar passages, see pp. 50 and 90).

On a number of occasions, however, G and M seem to forget the point and argue that static properties *can* count against modular theories. Here is an example:

> One test of whether a particular belief was the result of a module, an empirical generalization, or a theory might be to think about how we would react to an event that violated that belief. If our knowledge was really modular in a strict sense, we should not be able to represent the event at all, as our perceptual system can't represent the reality in a perceptual illusion and children can't represent the syntax of a pidgin language (GM, p. 79).

They then go on to argue (GM, pp. 80–81) that the fact that we can 'override' the 'perceptual representations' generated by our perceptual system in the case of illusions poses a problem for Spelke's modular 'core knowledge' account of our knowledge of objects. What's going on here? Since their argument at this point is not up to their usual standards of clarity, it's hard to be sure. But here's our guess. First of all, note the curious phrase 'modular in the strict sense'. What exactly does it mean? Though G and M never define it, the context suggests the following account: If a belief or 'perceptual representation' is the product of a body of knowledge which is 'modular in the strict sense' then that representation cannot be overridden by 'further

evidence'. To override a representation is to come to believe something other than what the representation claims to be the case. Thus if a strictly modular system generates a claim about some state of affairs, then we can't believe any alternative claim, indeed, perhaps we can't even conceive of an alternative. Now, with all this in place, they are in a position to argue that if some perceptual representation or belief can be overridden, then that representation cannot be the product of a system of knowledge that is 'modular in a strict sense'. So if 'in ordinary life there are many cases where our predictions about objects and disappearances would plainly have to override any innate core principles' (GM, p. 81), that spells trouble for the view that our knowledge about the behaviour of objects includes a system of core principles that is modular in the strict sense. Are there such cases in ordinary life? Of course there are:

> To use a simple and ubiquitous example, when we eat things, we do not assume that they will appear at the location where they disappeared. Instead, we assume they will be transformed or destroyed (GM, p. 81).

And thus it is that commonsense knowledge about the ultimate fate of what we eat refutes Spelke's theory of core knowledge.

Readers who have even the smallest sympathy with the principle of charity in interpretation are bound to feel rather uncomfortable with this argument against Spelke. Could it really be the case that one of the world's leading developmental psychologists has proposed a theory that can be refuted by the fact that most people know that what they eat is transformed into something else? The answer, not surprisingly, is no. Moreover, it is pretty clear where G and M's argument goes wrong. What they establish is that if our knowledge of object movements includes a core knowledge system of the sort that Spelke posits, then this core knowledge is not modular *in the strict sense*. Once we see what this means, however, it is clear that G and M are attacking a straw man. A body of knowledge is modular 'in the strict sense' only if we have to *believe* all the representations of the world that this body of knowledge generates. And Spelke makes no such claim about the representations of the world generated by core knowledge. She makes it clear that the core knowledge systems she posits are often response and task specific. Typically, the job of the representations they generate is not to determine what we believe but rather to guide some very specific activity. Thus, for example, Spelke maintains that different systems of core knowledge may be guiding visual tracking and predictive reaching, and under certain circumstances the predictions that these systems generate will be incompatible. (Spelke, 1994, p. 441; Carey and Spelke, 1996, p. 519) Obviously, we don't end up believing *both* predictions, and in some circumstances we won't end up believing either one. In determining what we believe, all

things considered, the pronouncements of core knowledge systems are eminently overrideable, and Spelke herself has been quite clear about this.[6]

4.3.2. Three Arguments Based on the Dynamic Properties of Theories Let us turn now to the second sort of evidence that might be invoked to determine whether the theory theory or a modularity theory does a better job at explaining the child's capacity in dealing with objects, agents and kinds. This, it will be recalled, is evidence about the dynamic properties predicted by the two sorts of theories. On the view that G and M urge (most of the time), it is this evidence that will be crucial.

> [Modular representations] will not have the dynamic features of theories. In particular, they will be indefeasible; they will not be changed or revised in response to evidence (GM, p. 50).
>
> The crucial evidence differentiating the two views lies in the dynamic properties of modules and theories, in how they develop (GM, p. 52).

What kinds of facts about the ways in which the representations or knowledge structures underlying children's abilities change over time do G and M think would count in favour of the theory theory and against modularity theories? As we read them they make three quite different suggestions.

The first suggestion: The child's representations change and become more accurate. The first suggestion is that the very fact that these representations do change, and that they generally get better or more accurate over time, counts in favour of the theory theory and against the modularity accounts.

> Modular representations do not lead to predictions through some set of inductive and deductive generalizations or through a process of theory testing, confirmation and disconfirmation. They lead to predictions because they are specifically designed by evolution to do so.
>
> A consequence of this is that modularity theories are, in an important sense, antidevelopmental. Apparent changes in representation occurring over time, on these views, can be accounted for only by processes outside the representational system itself. One

[6] Consider, for example, the following passage from Spelke et al., 1992, p. 629: 'These conceptions [of material objects] will be perpetuated over spontaneous development, because they serve to single out the objects about which humans gain knowledge ... *They can be overturned by instruction or disciplined reflection* if the student or scientist can use conceptions in a different domain of knowledge, such as mathematics, in order to single out a new set of entities in the physical world' (emphasis added).

> possibility is that they reflect the maturation of another innate structure, a later module coming on line . . .
>
> Performance deficits are also often invoked to deal with cases in which the child apparently has incorrect representations at one point, which are replaced by other more accurate representations later on. *Such sequences are predicted by the theory theory. They are anomalous, however, for modularity theories. It is easy to see why evolution might have designed a representational system that was inaccurate in some respects. It is much more difficult, however, to see why evolution would have designed a sequence of incorrect modules, each maturing only to be replaced by another* (GM, pp. 54–5; emphasis added).

Since this is not a theme that G and M pursue elsewhere in their book, we suspect that they don't really want to place much weight on the argument, and that's all to the good since the argument clearly won't bear much. It is, after all, a well established principle of developmental biology that *ontogeny often recapitulates phylogeny*.[7] Early in their development, primate fetuses have structures that resemble gills; as development proceeds, these structures disappear and are replaced by embryonic lungs. In this and many other standard examples, it appears that in the course of development organisms pass through stages that resemble the adult stage of their evolutionary forebears. Since this pattern is common in development, the modularity theorist is surely not going to be in the least embarrassed to propose that human children develop a sequence of incorrect modular representational systems each of which is replaced by a more accurate system as the child matures. These inaccurate systems may simply be the developmental echoes of adult systems in the organisms from which we evolved.

The second suggestion: If given the opportunity, children would learn about radically different worlds.

A second sort of fact which G and M claim would count in favour of the theory theory and against modularity theories would be a demonstration that children raised in a physical or psychological environment that is very different from ours end up with a correct theory about that environment, just as children raised in our environment end up with a correct theory about our environment.

---

[7] The biogenetic law, that ontogeny always recapitulates phylogeny, has been widely rejected in biology (Gould, 1977; Ridley 1993). There are numerous exceptions to the law. For instance, in some species, adults retain juvenile features of ancestral species (e.g., the Mexican axolotl, an aquatic salamander). In other species, there are developmental stages that probably don't recapitulate any ancestral stage (Ridley, 1993, p. 539). However, for our purposes the crucial point is simply that recapitulation is quite common. On this point, there is widespread agreement. For instance, after considering exceptions to the biogenetic law, Ridley writes, 'These exceptions not withstanding, recapitulation is noticeably common. Evolution has often proceeded by terminal addition' (Ridley, 1993, p. 539).

> There is, in principle, a simple experiment that could always discriminate modularity theory and theory theory. Place some children in a universe that is radically different from our own, keep them healthy and sane for a reasonably long period of time, and see what they come up with. If they come up with representations that are an accurate account of our universe, modularity is right. If they come up with representations that are an accurate account of their universe, the theory theory is right. Unfortunately, given the constraints of the federal budget, not to mention the constraints of conscience, this experiment is impossible (GM, p. 53).

This sort of 'in principle' crucial experiment is a recurrent theme in G and M's defence of the theory theory. On three occasions (pp. 81–2, 127–8 and 165), they elaborate on the proposal by describing aspects of the wonderfully weird worlds portrayed in the *Star Trek* series, and predicting that human children raised in those environments would develop accurate theories about them—theories very different from the ones that human children develop in our world. The *Star Trek* thought experiments are delightfully entertaining, so much so, indeed, that they led one of us to go out and rent the most recent *Star Trek* movie. (And much to his surprise, it is also delightfully entertaining!) However, it is our contention that in proposing these experiments as definitive (albeit practically impossible) ways to determine whether the theory theory or the modularity theory is correct, G and M are deeply confused.

A first problem with the proposal is the suggestion, made quite explicitly in the passage quoted above, that if modularity theory is correct, then all children who are sane and healthy will come up with the same theory (one that happens to be true of the world in which *we* live) no matter what world they inhabit. This might indeed be the case if the modularity theorist were committed to the claim that children are innately programmed to develop one specific theory in a given cognitive domain, in much the same way that they are innately programmed to develop a specific number of fingers or a specific configuration of major blood vessels. But, of course, modularity theorists need make no such claim. Chomsky's account of language acquisition is a paradigm case of what G and M would call a 'modularity theory'; indeed, it is an example that they frequently cite. Since there are estimated to be between 4,000 and 6,000 extant languages (Pinker, 1994, p. 232) and since there are obviously many more humanly possible languages that are no longer spoken or never have been, it is clearly quite wrong to suggest that modularity theories require that all sane and healthy children end up with the same theory regardless of the environment in which they are raised. G and M clearly recognize the problem, though they sometimes suggest that it's really just a minor difficulty since the number of theories a modular account could produce must be rather small.

> In some [modular] theories several alternative branching routes, so to speak, determine the eventual form the module may take. These are generally described as 'parameters' set by the input (Chomsky, 1986). Parameters allow for a somewhat richer developmental story than the one in which a module is simply turned on or off. The relation between the input and the setting of the parameter is still, however, a relation of triggering. In contrast, in a theory theory, by analogy with scientific theories, there should be indefinite scope for genuinely novel theories, not simply a choice of several options (GM, p. 55).

But a bit of elementary maths shows how odd it is to think that parameter-setting models only offer 'a choice of several options'. Suppose that in language, or in some other cognitive domain, there are 15 parameters that need to be set, and that each of these can take one of three values. On these very conservative assumptions, the module would make 3^{15} (or 14,348,907) options available. If we assume, not at all implausibly, that there are 25 parameters to set, the number of options rises to 847,288,609,443!

Once it is recognized that on modular accounts a child may have a huge number of options available, a much deeper problem with G and M's 'in principle' experiment comes into focus. There are imaginable outcomes of the experiment that might indeed refute the theory theory, but there are no outcomes that would refute the modularity theory. So, while the experiment might conceivably show that G and M's theory is *worse* than the competition, it couldn't possibly show that their theory is *better*. To see the point, let us imagine that we raise a child on Vulcan, a planet with an extremely bizarre *Star Trek*-inspired environment in which objects move, come into existence and go out of existence in ways that are radically different from the ways in which objects behave here on Earth. The experiment has two possible outcomes: either the child ultimately acquires a theory that accurately describes the principles governing the behaviour of objects in this bizarre world, or he does not. Neither of these outcomes poses a problem for the modularity theorist. If the child does acquire a correct theory, the modularity theorist can explain it in much the same way that she explains the child's acquisition of a language that is radically different from the language spoken by his biological parents. The language acquisition module makes millions of options available, and if the language spoken on Vulcan happens to be one of these, then the child will acquire Vulcan just as smoothly as children in France acquire French and deaf children (in the appropriate environment) acquire ASL. Similarly, a modularity theorist might maintain that the theory-of-appearance-module makes millions of theories available, and if an accurate theory about object movements on Vulcan happens to be one of them, then the child will acquire it. But, of course, even if the language acquisition module makes millions of options available to the child, there will also be endlessly many logically possible languages that the module does not make available. And if a child is exposed to one of these, he will not learn it,

though he may end up speaking some hitherto unspoken language, in much the same way that a child exposed to a pidgin ends up creating and speaking a creole. Analogously, if an accurate theory of object movements on Vulcan is not among the theories that the theory-of-appearance-module makes available, then the child will not acquire it. He may either acquire some other theory (the analog of acquiring a creole) or none at all. But whatever the outcome may be it will pose no problem for the modularity theorist, since her theory is compatible with all of the possible outcomes and does not predict any of them.

The situation is quite different for the theory theorist. If the child succeeds in acquiring an accurate theory of object movements on Vulcan, all is well. For the child, according to the theory theorist, is a little scientist equipped with 'particularly powerful and flexible cognitive devices, devices that are good at constructing accurate representations of new and unexpected worlds' (Ga, p. 490). So, if science can figure out the principles governing the behaviour of objects on Vulcan, the child should be able to do it too. But now suppose that the child *fails* to learn an accurate theory of Vulcan object movements. What account can the theory theorist offer in this case? Actually, there are a pair of sub-cases to consider. On G and M's account, the mechanisms underlying scientific reasoning in both the scientist and the child are powerful and flexible, but they are not completely unconstrained. 'The theory theory, after all, still assumes that not all the logically possible theories compatible with the evidence will actually be constructed. There are some possible theories that will be constructed by human beings, given a particular pattern of evidence, and some that will not' (GM, pp. 55–6). Thus some logically possible worlds will be so 'new and unexpected' that human science simply cannot construct accurate theories about them. If Vulcan turns out to be a world like this, then all is well for the theory theorist. The child can't figure out what's going on and neither can the scientist. But what about the other possible sub-case, the one in which human scientists *can* discover how things work on Vulcan but human children can't? Here, it would seem, the theory theorist is in trouble. For if the scientist and the child are using the same cognitive devices in constructing their theories, then it is puzzling that scientists can understand a world and children can't.

It is important to see that our argument, in the last two paragraphs, does not in any way depend on the claim that there are possible worlds whose laws or principles can be discovered by human scientists but not by human children. What we are arguing is that there are some outcomes of G and M's 'in principle simple experiment' that would be problematic for the theory theorist, while there are *no* outcomes that would be problematic for the modularity theorist. Thus, even if the experiment could be run, it could not possibly show that the theory theory is better than the modularity theory.

The third suggestion: evidence vs. triggering

The third strategy that G and M propose for showing that the theory theory is preferable to modularity theories focuses on the distinction

between developmental processes that are *triggered* by some environmental input and those that treat the environmental input as *evidence* and produce rational changes in internalized theories or other representations in response to this evidence. The distinction has played a very central role in debates between nativists and empiricists in cognitive science during the last several decades. Despite this, however, the distinction is hardly a model of clarity, and in many cases it is far from obvious whether a process counts as triggered or as rational and evidence driven. This is because those who invoke the distinction typically explain it by appeal to prototypical examples, and the further a process is from one of the standard prototypes, the harder it is to classify. Scientific reasoning is, of course, the favourite prototype of a rational, evidence-driven process. As a first pass at explaining the notion of triggering, many authors appeal to examples like imprinting. As Konrad Lorenz famously demonstrated, young goslings behave as though the first middle sized animate object that they see after hatching is their mother. Normally, of course, that animate object is their mother, thus it makes perfectly good evolutionary sense for them to be designed in this way. But there is nothing that much resembles evidence driven scientific reasoning going on in the process that leads a gosling to believe (or behave as if it believed) that the object in question is Mom. Rather, to use Fodor's memorable term, it is a 'brute causal process' (Fodor, 1981, p. 273) which when triggered by Lorenz rather than Mama Goose will lead the goslings to parade around after him and ignore their mother.

Now, as G and M correctly note, modularity theorists typically claim that the processes which lead to the availability and development of module-based knowledge structures are brute causal triggerings, not rational and evidence driven. Thus, if there is evidence that the processes driving the development of the theory of object appearances, the theory of action and the theory of kinds are rational and evidence driven rather than triggered, this will be a good reason to think that the theory theory is correct in these domains, and that modularity theories are false. But, as G and M are aware, a rational, evidence driven process 'may not be profoundly different from the case of a module with a great many parameters differently triggered by evidence' (GM, p. 56). Indeed, 'There is an interesting conceptual and formal question about whether a modular system with a sufficiently varied set of parameters and triggers would reduce to a [science-like] theorizing system, or vice versa' (GM, p. 55). Thus it is not surprising that in the final page or two of the chapters dealing with object appearances, actions and kinds, G and M admit quite candidly that they have no 'direct evidence' to offer for the claim that the developmental processes they've described are rational responses to evidence rather than merely being triggered.

> The theory theory proposes that the motivation for these changes comes from the infant's observations of the behavior of objects. It is the result of evidence. Yet again, we have no direct experimental

> support for this claim (GM, p. 185; for similar passages see p. 122 and p. 160)

In each case, G and M suggest that there may be some indirect evidence which shows that extensive exposure to relevant evidence, or making the relevant sorts of evidence more salient by encoding them linguistically, can accelerate the acquisition of the theory. But it is hard to see how any of this evidence could count in favour of the theory theory and against a modular theory that posits 'a great many different parameters differently triggered by evidence' (GM, p. 56).

This brings us to the end of our assessment of G and M's arguments for the claim that the theory theory is superior to the competition because it does a better job at explaining the available evidence. The conclusion for which we've been arguing is that G and M's arguments do not even come close to making their case. The argument which appeals to the static properties of theories is, by their own admission, entirely irrelevant since an advocate of a modularity account might well posit a mechanism which, when triggered, yields representations with all of those static properties. The three arguments based on the dynamic properties of theories fare no better. The fact that the child's representations of the world change and improve over time is not in the least embarrassing for the modularity theorist, since evolutionary processes have made parallel patterns of change a ubiquitous feature of development. Speculations about what children would learn in a radically different universe are doubly unconvincing. First, of course, there are no data. Second, no possible outcome of the hypothetical experiment could show that the modularity theory is wrong, though there are outcomes that would be hard for the theory theory to explain. Finally, as G and M admit, they have no direct evidence that the processes of cognitive development in early childhood are rational and evidence driven, as the theory theory requires, rather than triggered, as the modularity theorist would maintain. So the best that can be said for G and M's strong version of the theory theory is that it is (more or less) compatible with a broad range of evidence. But much the same can be said for modularity theories, and also for mixed accounts like the one favoured by Carey and Spelke in which modular core knowledge and science-like theory revision mechanisms both play a major role in cognitive development. Far from knocking out the competition, G and M have not even landed a solid blow.

5. Retro Philosophy of Science

As we noted in 4.1, many of the claims that G and M make about scientific theories are, and are intended to be, 'largely uncontroversial, not to say bland' (Ga, p. 496). They want their portrait of scientific theories to be 'compatible with many different philosophical accounts' (Ga, p. 495). But there is one cluster of questions about the underdetermination of theories by data

and about the role of convention and social factors in science on which G and M depart from this strategy and adopt what Gopnik describes as an 'unashamedly and self-consciously retro' position in the philosophy of science (Gb, p. 560). The position they adopt is quite an extreme one which forces them to reinterpret or reject much of what has been done in the philosophy, history and sociology of science during the last half-century. However, it is also a position which is, near enough, forced on them by their extreme, theories-all-the-way-down version of the theory theory.

Since the time of Duhem, in the early years of the twentieth century, it has been widely recognized in the philosophy of science that theories are underdetermined by their evidence in the sense that, from the point of view of deductive logic, any finite body of evidence is logically compatible with an indefinitely large range of theories. In the middle years of the century, Carnap and other logical empiricists devoted a great deal of effort to developing non-deductive logics which could narrow the underdetermination by assessing the degree to which a given hypothesis was supported by a body of evidence. On Carnap's account, however, the use of an inductive logic presupposes the choice of a language or a 'linguistic framework' which imposes significant constraints on the ontology of the theories set out in that framework. In Carnap's 'rational reconstruction' of scientific inquiry the process proceeds in two distinct stages: the first is the choice of a linguistic framework, the second is the formulation and testing of hypotheses and theories within that framework. The initial choice of a framework, and any subsequent decision to reject one framework and replace it with another 'cannot be judged as being either true or false because it is not an assertion. It can only be judged as more or less expedient, fruitful, conducive to the aim for which the language is intended' (Carnap, 1956, p. 207). Starting in the early 1950s, Quine mounted an enormously influential attack on Carnap's distinction between those parts of a theory that were mandated by the linguistic framework, and thus true by convention, and those that were left open by the framework and accepted or rejected on the basis of evidence. The distinction is tenable, Quine argued, only if we can draw a principled distinction between analytic and synthetic sentences, and there is no principled distinction to be found. So, while Carnap takes 'a pragmatic stand on the question of choosing between language forms [or] scientific frameworks'. Quine 'espouse[s] a more thorough pragmatism' (Quine, 1953, p. 46). 'Carnap maintains that ontological questions, and likewise questions of logical or mathematical principle, are questions not of fact but of choosing a convenient conceptual scheme or framework for science; and with this I agree only if the same be conceded for every scientific hypothesis' (Quine, 1966, p. 134).

As Gopnik sees it, all of this was a colossal mistake.

> The worm in the apple, the serpent in the garden of these accounts was the idea, which is in both Quine and Carnap, that the choice of the new language was 'conventional' or 'pragmatic.' The impli-

> cation, later made quite explicit in Quine, was that the decision was also arbitrary and simply socially determined . . . This idea set the stage for the later skepticism of Quine and the social constructionists who were to follow him (Ga, p. 503).

While there are many objections that might be raised about Gopnik's interpretation of Carnap and Quine as early advocates of the view that much of science is 'arbitrary and simply socially determined', these are really not to the point. For, whether or not Carnap and Quine held such views, there can be no dispute that in subsequent years many authors influenced by Quine,[8] and many others who probably never read a word of Quine, have argued that some aspects of theory change in science are indeed arbitrary or are determined by a process of social negotiation in which both political factors and the personalities of individual scientists play a role. There is an enormous literature of case studies aimed at showing how at one or another point in the history of science the way in which a theory evolved or was replaced by a radically different theory was decisively influenced by social, political and psychological factors, and also, at times, by chance.[9]

It is clear that if they are to maintain their strong version of the theory, G and M must reject all of these studies. For on their version of the theory theory the processes subserving theory change in science are identical to the processes subserving theory change in children—including very young children. But obviously the social and political processes which, according to these studies, play a substantive role in determining the content of new scientific theories in the course of theory revision are not processes that play any role in the development of theories in prelinguistic children. Nor do the arbitrary choices that putatively influence the development of scientific theories, since if there were such arbitrary choices in development, then we would expect different children to end up with significantly different theories about object appearances, actions and kinds rather than all ending up, as G and M claim, with essentially the same theories.[10] G and M can and do recognize that social factors play a role in facilitating science, by enabling the collection and dissemination of new evidence, for example. But their account of science as a 'spandrel, an epiphenomenon of childhood' (Ga, p. 490), can't allow these factors to play any significant role in determining the content of a new scientific theory which succeeds in replacing an old

---

[8] Including one of the present authors, see Stich, 1996, pp. 63–82.

[9] See, for example, Bloor, 1976; Feyerabend, 1981; Galison and Stump, 1996; Hull, 1990; Knorr-Cetina and Mulkay, 1982; Pickering 1992, and Shapin and Shaffer, 1985. For a helpful overview, see Downes, 1998.

[10] G and M do report data indicating developmental differences between Korean-speaking and English-speaking children, and these differences seem to correspond in interesting ways to linguistic differences between Korean and English. However, G and M do not argue that Korean and English speaking children develop substantively different theories. Rather, the differences are restricted to the rate at which certain abilities develop (GM, pp. 205–6).

one. On their view, the content of a new theory is entirely determined by the old theory, the evidence, and the innate processes of theory revision.

> The theory theory proposes that there are powerful cognitive processes that revise existing theories in response to evidence. If cognitive agents began with the same initial theory, tried to solve the same problems, and were presented with similar patterns of evidence over the same period of time they should, precisely, converge on the same theories at about the same time. (Ga, p. 494)

While it is clear that G and M's theory requires them to reject much of what has gone on in the history and sociology of science for the last few decades, we're not much impressed by the reasons they offer for doing so. As we read them, the principal reason they offer is that if claims about social factors or conventions or arbitrary choices playing a substantive role in theory change were correct, it would be difficult to explain why science leads to the truth.

> . . . social interaction, by itself, can't produce veridical theories or genuine theory change . . . (GM, p. 71)

> Assimilating all cognitive development to the model of socialization is . . . a dreadful mistake, allied to the dreadful mistake of postmodernism in general. The crucial fact about cognitive development, and cognition in general, is that it is veridical, it gives us a better understanding of the world outside ourselves. Purely social-constructivist views discount this fundamental link between the mind and the world. (GM, p. 72)

The reference to '*purely* social-constructivist views' is, we think, a bit of rhetorical overkill. For, while it is no doubt true that a pure social-constructivist view (whatever exactly that might be) would have a hard time explaining how science succeeds in producing veridical theories, it is no easier to see how science can produce veridical theories if social or political factors play *any* significant role in determining how theory change works in science. But this can hardly be taken to be a serious reason to reject these accounts of science, since as Gopnik herself acknowledges, it is 'profoundly mysterious' how *any* process of theory change 'generates representations that match up to the outside world' (Ga, p. 502). Since Gopnik provides no very persuasive justification for her 'self-consciously retro' taste in the philosophy of science, we think the link between those retro views and G and M's strong version of the theory theory constitutes a real liability for their theory. Perhaps all those studies putatively showing that social and political factors have played major roles in determining how scientific theories change are, like postmodernism, just a dreadful mistake. But it will be no easy task to show that they

are. And if even a modest number of those studies turn out to be correct, then G and M's version of the theory theory is simply mistaken.

Department of Philosophy
Rutgers University

Department of Philosophy
The College of Charleston

References

Bloor, D. 1976: *Knowledge and Social Imagery*. Chicago: University of Chicago Press.
Carey, S. and Spelke, E. 1996: Science and Core Knowledge. *Philosophy of Science*, 63, 515–33.
Carnap, R. 1956: Empiricism, Semantics and Ontology. In Carnap, R., *Meaning and Necessity*. Chicago: University of Chicago Press, 205–21.
Carruthers, P. and Smith, P. 1996: *Theories of Theories of Mind*. Cambridge University Press.
Chomsky, N. 1986: *Knowledge of Language: Its Nature, Origins, and Use*. New York: Praeger.
Davies, M. and Stone, T. (eds) 1995a: *Folk Psychology: The Theory of Mind Debate*. Oxford: Blackwell.
Davies, M. and Stone, T. (eds) 1995b: *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.
Deacon, T. 1997: *The Symbolic Species*. New York: W. W. Norton.
Downes, S. 1998: Constructivism. In *The Routledge Encyclopedia of Philosophy*. London: Routledge.
Feyerabend, P. 1981: *Realism, Rationalism and Scientific Method*. Cambridge University Press.
Fodor, J. 1981: The Present Status of the Innateness Controversy. In Fodor, J., *Representations*. Cambridge, MA: MIT Press.
Galison, P. and Stump, D. 1996: *The Disunity of Science*. Stanford: Stanford University Press.
Giere, R. 1996: The Scientist as Adult. *Philosophy of Science*, 63, 538–41.
Godfrey-Smith, P. 1991: *Teleonomy and the Philosophy of Mind*. PhD dissertation. University of California, San Diego.
Gopnik, A. 1984: Conceptual and Semantic Change in Scientists and Children: Why There Are No Semantic Universals. *Linguistics*, 20, 163–79.
Gopnik, A. 1988: Conceptual and Semantic Development as Theory Change: The Case of Object Permanence. *Mind and Language*, 3, 197–216.
Gopnik, A. 1996a: The Scientist as Child. *Philosophy of Science*, 63, 485–514.
Gopnik, A. 1996b: A Reply to Commentators. *Philosophy of Science*, 63, 552–61.
Gopnik, A. and Meltzoff, A. 1997: *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.
Gopnik, A. and Wellman, H. 1992: Why the Child's Theory of Mind Really Is a Theory. *Mind and Language*, 7, 145–71.
Gopnik, A. and Wellman, H. 1994: The Theory-Theory. In L. Hirschfeld and S.

Gelman (eds), *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York: Cambridge University Press, 257–93.
Gould, S. 1977: *Ontogeny and Phylogeny*. Cambridge, MA: Harvard University Press.
Hinton, G. and Nolan, S. 1987: How Learning Can Guide Evolution. In *Complex Systems*, vol 1. Technical report CMU-CS-86-128. Camegie-Mellon University, pp. 495–502.
Hull, D. 1990: *Science as a Process*. Chicago: University of Chicago Press.
Knorr-Cetina, K. and Mulkay, M. 1992: *Science in Context*. London: Sage.
Meltzoff, A. and Moore, M. 1983: Newborn Infants Imitate Adult Facial Gestures. *Child Development*, 54, 702–9.
Meltzoff, A. and Moore, M. 1989: Imitation in Newborn Infants: Exploring the Range of Gestures Imitated and the Underlying Mechanisms. *Developmental Psychology*, 25, 954–62.
Morton, A. 1980: *Frames of Mind: Constraints on the Common-Sense Conception of the Mental*. Oxford: Clarendon Press.
Pickering, A. 1984: *Constructing Quarks*. Chicago: University of Chicago Press.
Pickering, A. (ed.) 1992: *Science as Practice and Culture*. Chicago: Chicago University Press.
Pinker, S. 1994: *The Language Instinct*. New York: William Morrow and Co.
Potts, R. 1996: *Humanity's Descent: The Consequences of Ecological Instability*. New York: William Morrow.
Quine, W. 1953: Two Dogmas of Empiricism. In W. Quine, *From a Logical Point of View*. Cambridge, MA: Harvard University Press, 20–46.
Quine, W. 1966: On Carnap's Views on Ontology. In W. Quine, *The Ways of Paradox*. New York: Random House, 126–34.
Ridley, Mark. 1993: *Evolution*. Cambridge, MA: Blackwell Science.
Shapin, S. and Shaffer, S. 1985: *Leviathan and the Airpump*. Princeton: Princeton University Press.
Spelke, E. 1994: Initial Knowledge: Six Suggestions. *Cognition*, 50, 431–45.
Spelke, E., Breinlinger, K., Macomber, J. and Jacobson, K. 1992: Origins of Knowledge. *Psychological Review*, 99, 605–32.
Stich, S. 1983: *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
Stich, S. 1996: *Deconstructing the Mind*. Oxford University Press.
Stich, S. and Nichols, S. 1992: Folk Psychology: Simulation vs. Tacit Theory. *Mind and Language*, 7, 29–65.
Stich, S. and Nichols, S. 1995: Second Thoughts on Simulation. In Davies, M. and Stone, T. (eds), *Mental Simulation: Philosophical and Psychological Essays*. Oxford: Blackwell, 87–108.
Wellman, H. 1985: The Child's Theory of Mind: The Development of Conceptions of Cognition. In S. Yussen (ed.), *The Growth of Reflection in Children*. Orlando: Academic Press.
Wellman, H. 1990: *The Child's Theory of Mind*. Cambridge, MA: MIT Press.
Wellman, H. and Gelman, S. 1992: Cognitive Development: Foundational Theories of Core Domains. *Annual Review of Psychology*, 43, 337–375.