



The theory theory thrice over: the child as scientist, Superscientist or social institution?

Michael A. Bishop ^{a,*}, Stephen M. Downes ^b

^a *Department of Philosophy, 402 Catt Hall, Iowa State University, Ames, IA 50011-1306, USA*

^b *University of Utah Department of Philosophy, 260 Central Campus Drive Rm 341, Salt Lake City, UT 84112-9156, USA*

Received 3 January 2001; received in revised form 14 June 2001

Abstract

Alison Gopnik and Andrew Meltzoff have argued for a view they call the ‘theory theory’: theory change in science and children are similar. While their version of the theory theory has been criticized for depending on a number of disputed claims, we argue that there is a fundamental problem which is much more basic: the theory theory is multiply ambiguous. We show that it might be claiming that a similarity holds between theory change in children and (i) individual scientists, (ii) a rational reconstruction of a Superscientist, or (iii) the scientific community. We argue that (i) is false, (ii) is non-empirical (which is problematic since the theory theory is supposed to be a bold *empirical* hypothesis), and (iii) is either false or doesn’t make enough sense to have a truth-value. We conclude that the theory theory is an interesting failure. Its failure points the way to a full, empirical picture of scientific development, one that marries a concern with the social dynamics of science to a psychological theory of scientific cognition. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Scientific theory change; The theory theory; Children’s development; Gopnik and Meltzoff

In a recent series of publications, Alison Gopnik (1996a,b, 1999) and Gopnik and Andrew Meltzoff (1997) have argued for a bold empirical hypothesis they call the ‘theory theory’: theory change in science and in children are similar, and perhaps even identical. In defending the theory theory, Gopnik and Meltzoff are motivated

* Corresponding author.

E-mail address: mikebish@iastate.edu (M. A. Bishop).

by the idea (or the hope) that philosophers of science and developmental psychologists can learn from one another, that our respective fields might benefit from the other's insights. We applaud this motive and are on record defending the proposition that philosophers have a lot to learn from psychologists (see Bishop, 1992, 2000; Downes, 1993b). But in this particular case we believe that the attempt by Gopnik and Meltzoff to find deep cross-disciplinary insights runs into interesting and instructive problems.

A fundamental tenet of the theory theory is that 'specifying the parallels between cognitive development and science not only can help us to understand cognitive development, it also can help us to understand science itself' (Gopnik, 1996a, p. 486). But what does the word 'science' refer to in this passage? There are at least three different possibilities: an individual scientist, a rationally reconstructed 'Superscientist'¹ or the scientific community.² This three-fold ambiguity is not a harmless infelicity, a pedantic but eliminable worry that can be cleared up with a few scholastic distinctions. The ambiguity is a substantive one and leads to three quite different interpretations of the theory theory. The interpretations differ in terms of what process is likened to the conceptual development of children: the conceptual development of actual scientists, the conceptual development of a rationally reconstructed Superscientist, or (assuming this even makes sense) the conceptual development of the scientific community.

In the first three sections of this paper, we take up the three interpretations of the theory theory and argue that each has different kinds of problems. The child-scientist interpretation of the theory theory is false; the child-Superscientist interpretation is non-empirical (which is problematic since the theory theory is supposed to be a bold *empirical* hypothesis); and the child-community interpretation is either false or doesn't make enough sense to even have a truth-value. In Section 4, we argue that the theory theory goes wrong by not being bold enough. Gopnik and Meltzoff defend an embodied account of the nature and dynamics of scientific theories—an account that puts scientific theories in the heads of individual scientists. But they buy into the standard philosophical program of trying to account for disembodied, abstract facts about 'theory change' in science. This is the wrong theory for the job. We suggest that by abandoning the standard philosophical program, Gopnik and Meltzoff may well transform how we think about the nature and dynamics of scientific theory change.

¹ The Superscientist is a fictional construction that has none of the unfortunate limitations of brain-power, time or energy that vex real, human scientists. We take the Superscientist to be analogous to the Ideal Mathematician envisaged by Intuitionists in logic. The precise powers and characteristics of the Superscientist will depend on the philosophical predilections of the person who tells her story.

² We do not pretend that there is a clear demarcation about what counts as 'the' scientific community. Nor do we believe that the institution of science is best understood as a single scientific community. The expression 'the scientific community' here would require serious explication in order for the child-community interpretation to get off the ground. As we will argue in Section 3, however, the factors keeping this interpretation grounded are legion.

1. The child-scientist interpretation

The most natural way to understand the theory theory is in terms of a child-scientist analogy. This is how Gopnik and Meltzoff first describe the ‘central idea’ of the theory theory in *Words, thoughts and theories*: ‘the processes of cognitive development in children are similar to, indeed perhaps even identical with, the processes of cognitive development in scientists’ (Gopnik & Meltzoff, 1997, p. 3). Gopnik and Meltzoff follow a number of contemporary philosophers in defending a psychological view of theories. This view holds that a scientific theory is a particular set of cognitive representations and rules that transform those representations (Churchland, 1989; Giere, 1988; Kitcher, 1993).

We can think of a theory as a particular kind of system that assigns representations to inputs . . . The representations that it assigns are, however, distinctive in many ways . . . We can capture these distinctive structural features by talking about the specific abstract, coherent, causal, ontologically-committed, counterfactual supporting entities and laws of the theory . . . The representations are operated on by rules that lead to new representations . . . (Gopnik, 1996, p. 499; Gopnik & Meltzoff, 1997, pp. 42–44)

Given this conception of what a theory is, we can express the child-scientist analogy as follows. The structural, functional and dynamic properties of theories (i.e., of certain cognitive rules and representations) in particular children are similar (or identical) to the structural, functional and dynamic properties of theories (i.e., of certain cognitive rules and representations) in particular scientists. What exactly are these structural, functional and dynamic features of theories that are shared by children and scientists? In Section 1.1, we describe their structural and functional features. In Section 1.2, we describe the dynamic features of theories, and we argue that the theory theory’s account of the dynamics of theory change in children is not an accurate account of the dynamics of theory change for individual scientists. The child-scientist interpretation of the theory theory appears to be false. In Section 1.3, we consider the response Gopnik and Meltzoff offer to this objection. In Section 1.4, we consider the attempt to elude the division of labor problem by defending a weak version of the child-scientist interpretation of the theory theory. We admit that while some weak version of the theory theory may be true, it cannot do the philosophical work Gopnik and Meltzoff have set for it. Indeed, in Section 1.5, we note that Gopnik and Meltzoff take the theory theory to support a bold form of realism. We argue that the child-scientist interpretation of the theory theory, whether in its weak or strong form, provides no support for realism. So while some weak version of the theory theory may be true, it is not nearly as interesting philosophically as Gopnik and Meltzoff claim.

1.1. Theories in children and scientists: structural and functional features

On the child-scientist interpretation of the theory theory, the structural and functional features of the cognitive rules and representations that constitute theories in

the minds of individual children and individual scientists are similar (or perhaps identical). Gopnik and Meltzoff identify the following structural features of theories: theories contain abstract theoretical vocabulary (presumably in the Language of Thought), posit lawful relations, appeal to underlying causal structures, have ontological commitments, support counterfactuals, and are supposed to be true. Theories have certain functional features as well: they predict, and when a theory's prediction is violated, it produces surprise; they provide interpretations of evidence and define relevant and irrelevant evidence; theories also provide explanations, which Gopnik likens to orgasms (Gopnik, 1999)!

We will not challenge Gopnik's views about cognitive development in children, although others have (for example, Carey & Spelke, 1996; Harris, 1994). In fact, we plan to grant that the structural and functional features of theories are in the heads of individual scientists and have the features Gopnik identifies (despite misgivings—see Downes, 1992; Harris, 1994; Kusch, 1999). In describing the structural and functional features of theories, Gopnik is clearly pressing a child-*scientist* analogy. Gopnik says that violations of evidence produce surprise, and explanations are like orgasms. Surprise and orgasms happen to individual people—as opposed to groups or institutions. We readily grant that this way of understanding scientific theories is coherent and fruitful: a scientist's developing knowledge of a theory (her propositional knowledge) involves cognitive rules, representations and changes to those rules and representations.

1.2. Theory change in children and scientists: dynamic features

We have granted Gopnik's claim that the structural and functional features of theories in children and in scientists are identical.³ Now let us turn to theory change, to the dynamic features of theories. According to Gopnik and Meltzoff, a child will typically run through a five-stage sequence during theory change. The child (1) garners counter-evidence to the current theory, (2) denies that counter-evidence, (3) proposes *ad hoc* hypotheses (which undermine the coherence of the theory), (4) formulates a new theory, and (5) observes and experiments intensely to test the alternatives (Gopnik & Meltzoff, 1997, pp. 39–41; see also Gopnik, 1996a). Of course, these stages overlap in complex ways, but the important point for our purposes is that the individual Gopnikian child goes through *all* five stages in *each* episode of theory change. The child-scientist interpretation of the theory theory holds that the same is true of scientists: individual scientists go through *all* five stages in *each* episode of theory change.

It is important to distinguish the child-scientist interpretation of the theory theory from weaker (and more plausible) claims. The child-scientist interpretation of the theory theory does not merely say that for certain episodes of scientific theory change

³ In granting this we are leaving important theoretical and empirical difficulties to one side. For example, just exactly what cognitive structures are is a difficult question. And even if this were answered, how to ascertain empirically that the cognitive structures of children and scientists are the same would be very difficult.

some scientists garner counter-evidence to a dominant theory, or some scientists deny that evidence, or some scientists propose ad hoc hypotheses to support the dominant theory, or some scientists formulate new theories, or some scientists engage in intense observation and experimentation, or some scientists engage in more than one of these activities. Nor does the child-scientist interpretation of the theory theory say merely that the occasional scientist, working alone or in a small group, might develop an idiosyncratic theory and thereby personally undergo *all* five dynamic stages of Gopnikian theory change.

The child-scientist interpretation of the theory theory is committed to the view that *every time* a scientist adopts a new theory, that scientist goes through *all* five dynamic stages of theory change (garnering counter-evidence, denying that evidence, proposing ad hoc hypotheses, formulating new theories, and then intensely testing the alternatives). But this is false. The five-stage process Gopnikian children go through in episodes of theory change is not what a typical scientist—even a revolutionary scientist—goes through in an episode of theory change. Scientists who formulate new theories are often new to the field; so they typically aren't the same people who garner the evidence that disconfirms the entrenched theory (many have pointed this out, such as Hull, 1988; Kuhn, 1970). Nor do the scientists who garner counter-evidence always deny such findings (although some do pass through various stages of denial or incredulity). Certainly, the revolutionary scientists who build new theories that explain the counter-evidence don't always deny that counter-evidence (especially if they are newer to the field than the counter-evidence). Further, it is a rare and special scientist who can formulate a new scientific theory that explains the evidence that disconfirms an entrenched theory. Most scientists spend their lives articulating and applying theories they learn from their senior peers; the vast majority of working scientists never propose novel theories. In this century, there have been a lot more physicists than there have been new theories of physics. The same applies for chemistry, biology, geology and so on.

The basic point here is that the division of cognitive labor in science falsifies the child-scientist reading of the theory theory. Of course, Gopnik and Meltzoff are perfectly well aware of the division of cognitive labor in science (Gopnik, 1996a, pp. 486, 491–493; Gopnik & Meltzoff, 1997, pp. 13, 24–26). And they do not believe that it leads to a fundamental problem with their theory. We will now turn to their response to this problem.

1.3. *The theory theory's anti-social attitude*

Gopnik and Meltzoff offer three general lines of argument for their anti-social attitude toward the study of science. The first is that social structures aren't sufficient for explaining science's convergence on the truth. 'The socially oriented view of philosophy of science has always had a difficult time explaining how science gets it right at all. It has been difficult to reconcile with scientific realism . . .' (Gopnik, 1996a, p. 491). And again:

It is easy to see how the division of labor could result from the need for various kinds of evidence, and how that structure could lead to particular distinctive prob-

lems and patterns of timing in scientific change. The social hierarchy and the division of labor . . . may be genuinely helpful in solving certain problems. What is extremely hard to see, however, is how the hierarchy could lead to the truth in general, or how the division of labor could itself lead to theory formation or confirmation. (Gopnik, 1996a, p. 492)

It is true that social organization by itself does not explain how science manages to converge on the truth. But this doesn't mean that social organization is *irrelevant* to explaining how science gets it right (Hull, 1988; Kitcher, 1993). To see how misconceived this argument is, consider a perfectly analogous argument against the radical individualism implicit in the theory theory: the rich history of scientific thought proves that people having normal psychological organization does not by itself lead to scientific truth; therefore psychology is irrelevant to explaining how science gets it right. But of course, this argument is absurd. Science is ineliminably a psychological process, as it is ineliminably a social process.

Unfortunately, it seems that Gopnik and Meltzoff's anti-social attitude is driven in part by the fact that much contemporary sociology of science is informed by relativism. But nothing in principle dictates a relativistic sociology of science. Further, there is no good reason for the realist to reject the importance of social factors in an account of scientific theory change. From a realist's perspective, science is best seen as 'a social process in which scientists evaluate and criticize each other's work, leading to successive improvement' (Hull, 1988, pp. 361–362). In fact, a sociology of science open to realism might well provide part of a deep and interesting explanation for why science as a social institution populated by humans with certain psychological abilities has produced consensus on powerful, approximately true theories about the world (Kitcher, 1993).

The second line of argument offered for the theory theory's anti-social attitude grants that science is a social process but argues that the child-scientist analogy nonetheless holds true. Gopnik and Meltzoff argue that the child and the *scientist* are alike in being in a social environment: 'children are less isolated than the term "little scientist" is likely to imply. They live in a rich social structure with much opportunity for contradiction, instruction, and the linguistic transmission of information. We are not dealing with a contrast between a non-social process and a social one, but between two different types of social organization' (Gopnik, 1996a, p. 491; Gopnik & Meltzoff, 1997, p. 24). In fact, Gopnik and Meltzoff suggest 'that much of the social structure of science is an attempt to replicate the privileged sociological conditions of infancy' (Gopnik & Meltzoff, 1997, p. 25; Gopnik, 1996a, p. 493).

The third line of argument for their anti-social attitude toward science relies on the premise that scientists' basic cognitive equipment is not fundamentally altered by their social surround.⁴

⁴ Gopnik also attempts to defuse the division of labor problem by trying to explain it: 'It is characteristic of the child's problems that the evidence necessary to solve them is very easily and widely available . . . It is characteristic of scientific problems that the evidence necessary to solve them is rather difficult to obtain . . . It is this paucity of evidence that leads to the division of labor, and to many of the

Aside from the division of labor, the social hierarchy largely determines who will get the leisure and equipment to do cognitive work, and to whom other scientists should listen. The infant solves these problems without needing elaborate social arrangements. These are all differences between children and scientists, but again they do not imply differences in the fundamental cognitive processes that the two groups employ. (Gopnik, 1996a, p. 493)

Gopnik makes a similar point about the relative stability of cognitive processes across different social environments, by comparing scientists from different eras: ‘The institutional arrangements of Kepler or Newton or even Darwin were very different from those of contemporary scientists . . . However, it seems difficult to argue that the basic theory formation capacities of current scientists are strikingly superior to those of Kepler or Newton, in spite of the large differences in social organization’ (Gopnik & Meltzoff, 1997, p. 25; see also Gopnik, 1996a, p. 492).

These last two points—that the child and the scientist are alike in being in a social environment, and that scientists’ basic cognitive equipment is not fundamentally altered by their social surround—may be true, but they are orthogonal to the division of labor objection. The fundamental problem with the child-scientist interpretation of the theory theory is that it posits a dynamics of theory change that cannot be true of the vast majority of individual scientists given the division of cognitive labor in science. We can grant that children and scientists are both in social environments and that the basic features of human brains (and hence scientists’ brains) haven’t changed much over the past few centuries. But that doesn’t do anything to show that in episodes of scientific theory change, the entire, dynamic, five-stage process of theory change described by Gopnik and Meltzoff happens in the head of every normal individual scientist.

1.4. Weak versions of the child-scientist interpretation of the theory theory

To avoid the division of labor problem, it is very natural to weaken the child-scientist interpretation of the theory theory. Weak versions of the theory theory hold that the structural and functional features of theories in children and in scientists are similar, but the dynamical features of theories in children and scientists are not. A plausible and interesting version of the weak theory theory involves distinguishing between large-scale theory change (for example, Kuhnian revolutions) and small-scale theory change (for example, a scientist trying to perfect a pet theory).⁵ This weak interpretation retains the view that the dynamics of small-scale theory change

sociological institutions characteristic of science’ (Gopnik, 1996a, pp. 491–492). Even if this explanation is true, it is a red herring. The problem is that the dynamics of theory change in science are different from the dynamics of theory change in children because of the characteristic social features of science—*whatever the explanation of those characteristic social features of science may be.*

⁵ Alison Gopnik suggested this weak version of the theory theory to us (private correspondence), though Shaun Nichols (private correspondence) and Eric Schwitzgebel (private correspondence) also suggested variations on it.

is akin to the dynamics of theory change in children. But it grants that the dynamics of large-scale theory change are unlike the dynamics of theory change in children. However, there are various ways to soften this blow. One particularly interesting way is to argue that in large-scale theory change, social factors (such as publications, division of labor) permit scientists to experience the five stages of theory change vicariously.⁶

A number of obvious worries spring to mind. First, this view assumes a distinction between large-scale and small-scale theory change, and at the very least this requires some explication. Second, the notion of vicariously experiencing some stage of theory change is also in need of explication. How does one vicariously experience the discovery stages of theory change, such as someone else's formulation of a new theory or of ad hoc hypotheses, especially when this process of discovery is so often opaque, even to the discoverer? Is this process the same as vicariously experiencing the other stages of theory change (such as acquisition and denial of counter-evidence or testing alternative theories)? Does one vicariously experience someone else's scientific work by mentally rehearsing the steps they performed? Or does it merely require one to trust someone else's work? Or does it require something else? Third, the obvious possibilities for what counts as vicarious experience lead to problems for the theory theory. Any account of vicarious experience that demands significant pretense or mental rehearsal is likely to ruin the theory theory on historical grounds. For example, it is implausible in the extreme to suppose that scientists who are busy with their own labors have the time, energy or inclination to mentally rehearse all the scientific work they rely on. More often than not, they simply rely on work they deem trustworthy. On the other hand, if vicarious experience requires something very weak, like trusting others' work or judgments, then it is not clear why the theory theory is about children and *scientists*—as opposed to children and *any adult who follows scientific developments closely*.

These challenges to this version of the theory theory are serious, but are primarily questions of detail. And we don't propose to push them too hard, especially since there are other ways to weaken the child-scientist interpretation of the theory theory.⁷

⁶ This was suggested by Gopnik (private correspondence). It is reminiscent of a point made by Shapin and Schaffer (1985) about the way in which Boyle reported his air-pump experiments.

⁷ Another way one might articulate the weak theory theory is as follows: the *capacities* scientists and children bring to bear on theory change are similar. But we suspect this view is just too weak for Gopnik and Meltzoff. After all, the cognitive *capacities* scientists bring to bear on proposing and testing hypotheses about the fundamental nature of the universe are probably similar to the cognitive *capacities* philosophers bring to bear on proposing and testing new accounts of knowledge (or French chefs bring to bear on proposing and testing new recipes).

Here is yet a third way to articulate a weak version of the theory theory: children go through a five-stage process of theory change, but few (if any) scientists go through that entire process during theory change in science. Rather, the theories of some scientists change in ways that are analogous to some elements of the five-stage process of theory change in children. For example, there are similarities between the cognitive dynamics in adult scientists and children when they garner counter-evidence to a received theory, when they deny that counter-evidence, when they propose ad hoc hypotheses to save the received theory, when they formulate new theories, and when they intensely test various theoretical alternatives. But this version of the theory theory does not try to shoehorn all aspects of theory change into the heads

We will grant that some weak version of the theory theory might well avoid the division of labor problem, and might even be empirically plausible. But from the perspective of Gopnik and Meltzoff, a serious problem faces this view. The child-scientist interpretation of the theory theory provides no support to scientific realism. Gopnik and Meltzoff want to claim not only that the theory theory is true, but that it is also philosophically interesting—because it supports scientific realism. In the following section, we will argue that the child-scientist interpretation of the theory theory, in both its strong and weak forms, does not support scientific realism.

1.5. Realism and the child-scientist interpretation of the theory theory

Gopnik and Meltzoff are die-hard realists. They believe—or at least claim to have ‘faith’ in the view (Gopnik & Meltzoff, 1997, p. 47)—that theory change in science tends to converge on the truth. Gopnik says, ‘I can imagine that I might come to doubt the idea that science gets at the truth about the world, as clergymen in the nineteenth century came to doubt the ontological underpinnings of theology, but then I would stop doing science, as honest clergymen quit the church’ (Gopnik, 1996b, p. 555). We intend to grant for the sake of argument that our current theories are roughly true and that the history of science is, by and large, a history of ever closer approximations to the truth.⁸

Given the reasonable assumption that children’s theories of physics, mind and so on tend to get closer to the truth over time, it appears that the theory theory may win the day on this issue: theory change in children and in science are the same in that they both converge on the truth. Further, Gopnik and Meltzoff have a hypothesis that explains why children’s theories and scientists’ theories both converge on the truth.

The assumption of cognitive science is that human beings are endowed by evolution with a wide variety of devices—some quite substantive and domain-specific, others much more general and multipurpose—that enable us to arrive at a roughly veridical view of the world . . . A cognitive view of science, and in particular, a view that identifies cognitive change in science and childhood, might provide at least a partial explanation of the most important thing about science, namely that it gets things right. (Gopnik & Meltzoff, 1997, pp. 15–16)

Let us examine this claim. In this passage, Gopnik and Meltzoff take seriously the idea that the theory theory can provide a partial, unified explanation for truth-conver-

of individual scientists.

Whatever their relative merits, these weak versions of the theory theory, like the version described in the text of this paper, are not much help to Gopnik and Meltzoff. First, their texts give no indication that they hold any of these weak views. Second, the weak views do not amount to the requisite bold hypothesis. Finally, it is necessary for any plausible interpretation of their view that it imply realism; the weak views fail to meet this criterion.

⁸ For sophisticated dissenting views, see van Fraassen (1980), Laudan (1984) and Fine (1984).

gence in children and in science. But if this is right, then they can't have in mind the *child-scientist* interpretation of the theory theory. Let's assume that a normal child's theories tend to converge on the truth.⁹ But the same is not true for every normal scientist. Gopnik and Meltzoff would surely agree that many individual scientists spend their entire careers going down the wrong path, defending and articulating false theories that don't converge on the truth. *Science* may well 'get things right', but *individual scientists* often don't. Taken as an empirical hypothesis about individual cognition, the theory theory should account for the cognizing of scientists who get it dead right and the scientists who get it dead wrong. How can there be anything in the resulting descriptions of these cognizers that implies realism? The child-scientist interpretation of the theory theory simply does not provide any support for scientific realism.¹⁰

Another way to see that the child-scientist interpretation of the theory theory provides no aid or comfort to the realist is to note that it can be co-opted by a rabid anti-realist. For example, the anti-realist might grant that Gopnik and Meltzoff have accurately described the psychological mechanisms individual scientists use when doing science. But the anti-realist might nonetheless argue that our best current scientific theories are very probably false. For example, the anti-realist might grant that scientists (actually or vicariously) (1) garner counter-evidence to a current theory, (2) deny that counter-evidence, (3) propose ad hoc hypotheses (which undermine the coherence of the theory), (4) formulate a new theory, and (5) observe and experiment intensely to test the alternatives. Nothing in the theory theory's description of the scientist's cognizing tells us whether she is involved in a progressive episode (moving closer to the truth) or not (moving down a scientific cul-de-sac and away from the truth). Thus the anti-realist could argue for anti-realism, on fairly standard grounds. For example, she might argue inductively that since past successful theories have been shown to be false, our current successful theories are probably false as well (see Laudan, 1984).

Of course, we are not claiming that any particular argument succeeds in showing that anti-realism is probably true. Rather, our claim is that one could consistently make such a case and gladly embrace the child-scientist interpretation of the theory theory. The child-scientist interpretation of the theory theory is silent about the realism-anti-realism controversy. It tells us only about what happens in the heads of scientists. It does not tell us anything about what (non-psychological) entities the universe contains nor about how such entities behave. But scientific realism is essentially the thesis that the universe contains most of the central entities posited by our best theories and that those entities behave in roughly the ways depicted by those

⁹ There is evidence that children do not always replace theories with other theories that are closer to the truth (Harris, 1994). Still, we can assume that, viewed over a suitable period of time, every normal child's theories do tend to converge on the truth. As we argue in the text, the same is not true for every normal scientist.

¹⁰ Gopnik and Meltzoff lend credence to our point as, when discussing realism, they typically say things such as that a 'cognitive view of *science*' (not scientists) might provide a partial explanation for why '*science*' (not scientists) 'gets things right' (Gopnik & Meltzoff, 1997, p. 16; emphasis added).

theories. The theory theory, if it is to support Gopnik and Meltzoff's bold realism, must be about more than the workings of individual minds. For if the history of philosophy proves anything, it proves that it is difficult to draw conclusions about the mechanics of the universe from claims about the mechanics of the mind. There are no easy moves from beliefs to bosons.

2. The child-Superscientist interpretation

We have argued that the (strong version of the) child-scientist interpretation of the theory theory is falsified by the division of cognitive labor in science. Given that Gopnik and Meltzoff know about the division of cognitive labor in science (and in fact practise it), generosity demands that we look for another interpretation of the theory theory, one that avoids this objection. Further, Gopnik and Meltzoff believe that the theory theory bolsters a bold scientific realism. But the child-scientist interpretation does not do this. So once again, we are moved to look for another interpretation of the theory theory. We will suggest that sometimes Gopnik and Meltzoff are proposing (or at least assuming) a child-Superscientist interpretation of the theory theory.

On the child-Superscientist interpretation, the structural, functional and dynamic properties of theories (i.e., of certain cognitive rules and representations) in particular children are similar (or identical) to the structural, functional and dynamic properties of theories (i.e., of certain cognitive rules and representations) in a rationally reconstructed Superscientist. While this interpretation may appear *prima facie* wild and ungenerous, it has two significant virtues as a reading of Gopnik and Meltzoff's texts.

The first virtue of this interpretation is that it fits with the 'retro'¹¹ philosophy of science that Gopnik and Meltzoff embrace (Gopnik, 1996b, p. 590). Gopnik and Meltzoff admire the logical empiricists (Gopnik & Meltzoff, 1997, p. 33). Logical empiricists, as well as other 'retro' philosophers such as Popper and Lakatos, advocated the development of an account of scientific reasoning that idealized away from the day-to-day practices of individual scientists. When Gopnik and Meltzoff discuss the dynamics of theory change in science—the process and its tendency toward the truth—they abandon the cognitive apparatus of the individual scientist. There is the occasional reference to the very best work of the very greatest scientists. But it is their achievements that interest Gopnik and Meltzoff, not the messy details of the processes by which the Galileos, Newtons and Darwins arrived at their results. When discussing the dynamics of scientific theory change, Gopnik and Meltzoff seem engaged in rational reconstruction: idealizing away from the rough-and-tumble of science and conceiving of it as if it had been done by a single individual, a Superscientist, who has none of the standard cognitive limitations or motivational 'impurities' of real scientists.

The second virtue of the child-Superscientist interpretation of the theory theory

¹¹ 'Retro' is Gopnik and Meltzoff's term, not ours.

is that it neatly avoids the two problems that plagued the *child-scientist* interpretation of the theory theory. The division of cognitive labor is not a problem for the rationally reconstructed Superscientist. She does it all: she gathers the counter-evidence to the dominant theory, denies that counter-evidence, tries to fix up the dominant theory by proposing ad hoc hypotheses, develops new theories, and tests the alternatives. Nor is relativism a threat for the child-Superscientist interpretation. Recall, the child-scientist interpretation seemed to imply that all normal scientists converge on the truth. But given the history of dead-ends in science, this can imply a kind of relativism that is clearly anathema to Gopnik and Meltzoff. But if we imagine science as having been done by a single Superscientist, then it makes sense to suppose that she has tended to converge on the truth over time.

The child-Superscientist interpretation is practically forced upon us at this point. The child-scientist interpretation of the theory theory is false; Gopnik and Meltzoff are perfectly aware of the fact that disconfirms it; and they sometimes seem to have some other interpretation in mind (i.e., when discussing realism). Given that Gopnik and Meltzoff adopt a fiercely psychological view of the nature of theories, where are we to locate the psychological entities that constitute theories if not in the minds of individual scientists? Given the ‘retro’ philosophy of science defended by Gopnik and Meltzoff, a natural answer seems to be: in the mind of a rationally reconstructed Superscientist.

A rational reconstruction is not the story of how science actually happened, but how it ‘should’ have happened (Lakatos, 1970). It is based on a (perhaps implicit) normative conception of scientific rationality. It is a Stalinesque approach to the history of science. Inconvenient historical events and embarrassments that don’t fit with the relevant ideology are eliminated from the official history. We expunge Kepler’s harmonies, Darwin’s ideas about heredity, Newton’s alchemy, God’s place in Newton’s physics (What keeps gravity from making everything in Newton’s universe smash together? The hand of God.), and so on. Historical doubters and hold-outs are struck from the record (only those who were wrong, of course). Theory change happens when, from a precisely characterized perspective, it ‘ought’ to have happened. And if our rational reconstruction has to ignore certain evidence historical figures had, or attribute to them evidence they didn’t have, all the while polishing up their evidence with contemporary knowledge, so much the better.

A rational reconstruction of the history of science is a radical idealization. There is nothing wrong with radical idealizations in general. They have been used to good effect in science and philosophy. But idealizations can have various degrees of reality in them; some idealizations model quite closely some aspect of the world, while others are wild fictions. An idealization that takes the extraordinarily rich and complex history of science and shoehorns it into the mind of a single Superscientist might be useful for some purposes, but as a model of a real-world dynamical process, it is science fiction. So if the theory theory is the claim that there is an analogy or identity between theory change in children and theory change in a *rationally reconstructed Superscientist*, then it is no longer a bold *empirical* hypothesis at all. Instead of being an empirical claim that says that two real-world processes are alike, this version of the theory theory says that some actual process is like some non-actual

process. We have no empirical methods that allow us to investigate how non-actual Superscientists think or how their thinking is related to that of children. One of Gopnik and Meltzoff's avowed aims is to show that the theory theory is 'more than a vague metaphor' (Gopnik & Meltzoff, 1997, p. 14). It is hard to see how they could have succeeded in this aim if their intention is to defend the child-Superscientist interpretation of the theory theory.

3. Socializing the theory theory?

The theory theory claims that there is a similarity (or identity) in the structure and dynamics of theory change in individual children and in something else. But we're having some trouble identifying what that something else might be. So far we have argued that it is not individual scientists, and it is not rationally reconstructed Superscientists. There is another possibility, which is that the theory theory posits a likeness between the conceptual development of children and (something like) the conceptual development of a community of scientists. We suspect that Gopnik and Meltzoff are not in the least bit tempted by this interpretation, and for good reason. For one thing, while it makes sense to talk about the conceptual development of a child, it is unclear how to make sense of the conceptual development of a community (as opposed to their individual members). But even putting this problem aside, there are many features of the social process of science that are simply left out of the Gopnik–Meltzoff account of theory change in children. For example, it is an essential part of the scientific story that there are many individuals, each with their own psychological apparatus (see Solomon, 1994). Different scientists often have different understandings of the theories under debate, including misunderstandings, and very different attitudes toward the theories (faith, belief, willingness to consider, agnosticism, doubt, disbelief, hate, and so on).¹² This cacophony of understandings and attitudes is essential to the process. We simply can't explain the vicissitudes of (say) Darwin's theory since the mid-nineteenth century without noting that lots of people had lots of different attitudes toward various different conceptions of it (Bowler, 1988). Further, the social process consists of very complex networks of social relations (such as relations of power and trust, political and social dynamics, personal, institutional and national loyalties, and so on) and a very complex technological backdrop, whose entirety may not be within any single person's ken. Whatever the elements and dynamics of these social processes, they will be markedly different from the cognitive processes of a single individual (see Downes, 1993a). While no complete account of any community, religious, economic or scientific, can ignore psychological factors, it is doubtful that such an account could be reduced entirely to psychological factors. And even if it could be, it couldn't be reduced to the limited psychological factors Gopnik and Meltzoff cite in their descriptions of

¹² Giere (1988) puts this discussion in terms of scientists' various interests, cognitive or otherwise.

theory change. Children's cognitive development cannot serve as a model for how an epistemic community operates.

4. Conclusion

Gopnik and Meltzoff follow a long tradition in science studies of researchers who are not clear about their unit of analysis. Ron Giere called attention to this failing over a decade ago (1988). For years, philosophers, historians and sociologists have talked about the dynamics of theory change in 'science'. It is by now standard practice to assume that any acceptable account of or proposition about scientific theory change must account for certain standard case studies. These include case studies of superstar scientists (such as Galileo, Newton, Darwin, Einstein) as well as case studies of superstar theories (such as Copernicanism, Newtonian mechanics, wave theories of light, Darwinism, quantum mechanics, the theory of relativity). If a view about scientific theory change is supposed to be able to account for theory change in all these cases, then the unit of analysis issue inexorably arises: what is a view about theory change *about* if it is supposed to account for case studies of individual scientists as well as case studies of scientific communities (for example, Darwinists)? Such views are typically thought to be true of abstractly characterized theories. In recent years, philosophers have offered a number of different accounts of what theories are. Some have identified them with abstract mathematical objects (Giere, 1988; Lloyd, 1988; Suppe, 1977; van Fraassen, 1980). Others have identified theories with sets of propositions (Kitcher, 1993). On both accounts of theories, we can talk about a scientific theory and changes to a scientific theory without necessarily talking about any person who might have held that theory. Let us call these disembodied views of scientific theories.

A disembodied view of theories allows philosophers to get away with not being clear about their units of analysis. By discussing theory change abstractly, philosophers can support their accounts by appealing to very different kinds of evidence. Sometimes what happens to an individual scientist, or a 'typical' scientist, or a 'good' scientist, or an idealized scientist might be taken to support an account of theory change. At other times, how the dominant opinions of a community of scientists change over time is taken to support an account of theory change. And on occasion such an account is thought to be supported by the fact that it fits with a historical reconstruction of the important ideas that arose among a community of scientists. We have serious reservations about whether any single view of theory change can reasonably account for all these different phenomena, but we will not pursue these doubts here.

The fundamental problem with the theory theory is that it is an *embodied* account of scientific theory change that tries to account for *disembodied* 'facts' about theory change in science. Sometimes, especially when defending realism, Gopnik and Meltzoff are trying to account for disembodied 'theory change' in science. But at other times, especially when likening theory change in scientists to that in children, they are saying that theory change involves real, cognitive processes, embodied in

particular individuals. Their mistake is in trying to account for disembodied ‘theory change’ in science with an embodied account of theories. The inevitable result is that they try to cram a multitude of variegated and disembodied facts into limited and embodied theories, and so end up shoehorning too much into the heads of individual scientists.

We have argued that important aspects of the theory theory do not stand up to scrutiny. Our analysis raises the strong suspicion that Gopnik and Meltzoff are beholden to disembodied accounts of theory change that are philosophical vestiges of our non-naturalistic past. Rather than embrace these vestiges, we believe that Gopnik and Meltzoff should abandon them. A full, empirical picture of scientific development will require theories about the social dynamics of science as well as theories about individual scientists’ cognitive processes. Developmental psychology is a potentially rich resource for insights about this latter project. But Gopnik and Meltzoff should reject as irrelevant appeals to abstract, disembodied accounts of theory change, which could only be instantiated by a Superscientist. Gopnik and Meltzoff might well be on the cusp of transforming our understanding of theory change in science, but they can only do this successfully by sticking to bold, clear empirical hypotheses.

Acknowledgements

We would like to thank Alison Gopnik, Rebekah Miner, Ram Neta, Shaun Nichols, Eric Schwitzgebel, Cindy Stark and several anonymous referees for very helpful comments on earlier drafts of this paper.

References

- Bishop, M. (1992). The possibility of conceptual clarity in philosophy. *American Philosophical Quarterly*, 29, 267–277.
- Bishop, M. (2000). In praise of epistemic responsibility: How lazy and ignorant can you be? *Synthese*, 122, 179–208.
- Bowler, P. J. (1988). *The non-Darwinian revolution: Reinterpreting a historical myth*. Baltimore, MD: Johns Hopkins University Press.
- Carey, S., & Spelke, E. (1996). Science and core knowledge. *Philosophy of Science*, 63, 515–533.
- Churchland, P. (1989). *A neurocomputational perspective: The nature of mind and the structure of science*. Cambridge, MA: MIT Press.
- Downes, S. M. (1992). The importance of models in theorizing: A deflationary semantic view. In D. Hull, M. Forbes & K. Okruhlik (Eds.), *Proceedings of the philosophy of science association* (Vol. 1, pp. 142–153). East Lansing, MI: Philosophy of Science Association.
- Downes, S. M. (1993a). Socializing naturalized philosophy of science. *Philosophy of Science*, 60, 452–468.
- Downes, S. M. (1993b). Modeling scientific practice: Paul Thagard’s computational approach. *New Ideas in Psychology*, 11, 229–243.
- Fine, A. (1984). The natural ontological attitude. In J. Leplin (Ed.), *Scientific realism* (pp. 83–107). Berkeley: University of California Press.
- Giere, R. (1988). *Explaining science*. Chicago: University of Chicago Press.

- Gopnik, A. (1996a). The scientist as child. *Philosophy of Science*, 63, 485–514.
- Gopnik, A. (1996b). A reply to commentators. *Philosophy of Science*, 63, 552–561.
- Gopnik, A. (1999). Explanation as orgasm. *Minds and Machines*, 8, 101–118.
- Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts and theories*. Cambridge, MA: MIT Press.
- Harris, P. L. (1994). Thinking by children and scientists: False analogies and neglected similarities. In L. A. Hirshfield & S. A. Goldman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 294–315). New York: Cambridge University Press.
- Hull, D. (1988). *Science as a process*. Chicago: University of Chicago Press.
- Kitcher, P. (1993). *The advancement of science*. Oxford: Oxford University Press.
- Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
- Kusch, M. (1999). *Psychological knowledge: A social history and philosophy*. London: Routledge.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge: Cambridge University Press.
- Laudan, L. (1984). A confutation of convergent realism. In J. Leplin (Ed.), *Scientific realism* (pp. 218–249). Berkeley: University of California Press.
- Lloyd, E. (1988). *The structure and confirmation of scientific theories*. Princeton: Princeton University Press.
- Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle and the experimental life*. Princeton: Princeton University Press.
- Solomon M. (1994). Multivariate models of scientific change. In D. Hull, M. Forbes & R. M. Burian (Eds.), *Proceedings of the Philosophy of Science Association* (Vol. 2, pp. 287–297). East Lansing, MI: Philosophy of Science Association.
- Suppe, F. (1977). *The structure of scientific theories*. Urbana, IL: University of Illinois Press.
- Van Fraassen, B. (1980). *The scientific image*. Oxford: Clarendon Press.