Dædalus

Journal of the American Academy of Arts & Sciences

Winter 2004

Alison Gopnik

Finding our inner scientist

In 1946, the philosopher of science Karl Popper had a fateful meeting with the philosopher of language Ludwig Wittgenstein at the Cambridge Philosophy Club. In a talk to the Club, with Wittgenstein in the audience, Popper described several "philosophical problems" – important, difficult questions that he thought would one day be answered. Here Popper was issuing a direct challenge to Wittgenstein, who had argued that philosophy could only analyze linguistic puzzles – not solve any real problems.

The visit has become most famous for the subsequent controversy among eyewitnesses over whether or not Wittgenstein's response to this challenge was to angrily brandish a fireplace poker at Popper.

But there is a more interesting aspect to the story. One of the problems Popper described was the problem of causal induction: How is it possible for us to correctly infer the causal structure of the world from our limited and fragmentary experience? Popper claimed that this problem would one day be solved, and he turned out to be right. Surprisingly, at least part of the solution to the problem comes from a source about as far removed from the chilly Cambridge seminar room of fifty years ago as possible – it comes from babies and young children.

The past thirty years have been a golden age for the study of cognitive development. We've learned more about what babies and young children know, and when they know it, than we did in the preceding two thousand years. And this new science has completely overturned traditional ideas about what children are like.

The conventional wisdom, from Locke to Freud and Piaget, had been that babies and young children are irrational, egocentric, pre-causal, and solipsistic, governed by sensation rather than reason, and impulse rather than intention. In contrast, the last thirty years of research have taught us that even the youngest infants – literally newborns – already know a great deal about a wide range of subjects. Moreover, we have been able to chart consistent changes

Alison Gopnik, professor of psychology at the University of California at Berkeley, conducts research on the ways that children come to understand the world around them. She has written many articles and monographs and is the coauthor (with Andrew N. Meltzoff and Patricia K. Kuhl) of "The Scientist in the Crib: What Early Learning Tells Us About the Mind" (1999).

in children's knowledge of the world as they grow older. Those changes suggest that even the youngest babies are solving Popper's problem: somehow they accurately learn about the causal structure of the world from their experience.

Consider how children come to understand one particularly important aspect of the world – the fact that other people have emotions, desires, and beliefs and that those mental states cause their behavior. All of us know that other people have minds in spite of the fact that we only see the movements of their physical bodies. This raises another ancient philosophical question: How do we come to know other minds?

In the last fifteen years, a great deal of empirical research has begun to illuminate the intuitive psychology of even the youngest human beings. Infants seem to be born believing that people are special and that there are links between their own internal feelings and the internal feelings of others. For example, newborns can imitate facial expressions: when an experimenter sticks his tongue out at the baby, the baby will stick out her own tongue; when he opens his mouth, she will open hers; and so on. In order to do this, newborns must be able to link their own internal kinesthetic sensations, the way their mouth feels from the inside, to the facial gestures of another person – that pink thing moving back and forth in the oval in front of them.[1]

By a year, babies seem to understand that mental states can be caused by external objects. For example, fourteen-month-olds saw an experimenter make a disgusted face as she looked inside one box, and a happy face when she looked inside another box. Then she gave the children the boxes. The children cheerfully opened the 'happy' box but kept the 'disgusted' box shut.[2] In another experiment, infants seemed to predict that a hand that had reached toward an object would continue to reach toward it even when it was placed at a new location – just as their own hands would. (They did not, however, make this same prediction about a stick that had made contact with an object.)[3]

By two, children seem to understand that their own desires may differ from the desires of others. And by two and a half, they extend this understanding to perception. In one study, the experimenter demonstrated disgust toward a food that the baby liked (goldfish crackers) and happiness toward a food that the baby did not like (raw broccoli), and then asked the baby to "give [her] some." Fourteen-month-olds always gave her the crackers, but eighteen-month-olds gave her the broccoli.[4] In another experiment, thirty-month-old children could accurately predict that someone on one side of an opaque screen would see a toy placed there, but someone on the other side of the screen would not.[5]

[1] Andrew N. Meltzoff and Wolfgang Prinz, eds., *The Imitative Mind: Development, Evolution, and Brain Bases* (New York: Cambridge University Press, 2002).

[2] Betty M. Repacholi, "Infants' Use of Attentional Cues to Identify the Referent of Another Person's Emotional Expression," *Developmental Psychology* 34 (5) (September 1998): 1017 – 1025.

[3] Amanda L. Woodward, Jessica A. Sommerville, and Jose J. Guajardo, "How Infants Make Sense of Intentional Action," in Bertram F. Malle and Louis J. Moses, eds., *Intentions and Intentionality: Foundations of Social Cognition* (Cambridge, Mass.: MIT Press, 2001), 149 – 169.

[4] Betty M. Repacholi and Alison Gopnik, "Early Reasoning About Desires: Evidence from 14- and 18-Month-Olds," *Developmental Psychology* 33 (1) (January 1997): 12 – 21.

[5] John H. Flavell, Barbara A. Everett, and Karen Croft, "Young Children's Knowledge

By four, children can understand that beliefs, as well as desires and perceptions, may differ, and that beliefs may be false. For example, you can show children this age a candy box that, much to their surprise, turns out to be full of pencils. Three-year-olds will say that they always thought that there were pencils in the box, and that everyone else will think that there are pencils inside, too. But four-year-olds understand that they and others may falsely believe that there are candies in the box.[6]

By six, children start to understand that beliefs may be the result of interpretation, and that different people may interpret the world differently. When you give five-year-olds a small glimpse of a picture – a triangular fragment that might imply a sailboat, or a witch's hat, or many other things – they don't understand at first that people might interpret this fragment in different ways. But by six or so they get this right.[7]

At each point in development children know some quite abstract and sophisticated things about how the mind works, knowledge that leads them to surprisingly accurate and wide-ranging predictions and explanations. They seem to understand something about how events in the world cause different mental states, and about the way these mental states in turn cause particular human actions. Yet they fail to understand other aspects of the causal structure of mental life – misunderstandings that lead to surprisingly inaccurate but consistent predictions and explanations. As they get older, the misconceptions fade away and their causal knowledge becomes more extensive and precise.

Evidence seems to play an important role in these developments. For example, younger siblings from large families, who have a lot of experience with a variety of other minds, develop this understanding more quickly than solitary only children.[8] We can also show that giving young children relevant evidence can actually accelerate their developing understanding of the mind. For example, we can, shades of Popper, set out to show children who do not yet understand false beliefs that their predictions about another person's actions can be systematically falsified; we can show them that someone who sees the closed box will, in fact, say there are candies inside of it. A month later, children who saw evidence that they were wrong were more likely to understand how false beliefs really work than children who did not.[9]

We can tell very similar stories about children's developing causal knowledge of everyday physical phenomena, like gravity and movement, and everyday biological phenomena, like illness and growth. These patterns of development have led many of us to draw an analogy between children's learning and the historical development of scientific theo-

about Visual Perception: Further Evidence for the Level 1 – Level 2 Distinction," *Developmental Psychology* 17 (1) (January 1981): 99 – 110.

[6] Josef Perner, Susan R. Leekam, and Heinz Wimmer, "Three-Year-Olds' Difficulty with False Belief: The Case for a Conceptual Deficit," *British Journal of Developmental Psychology* 5 (2) (June 1987): 125 – 137.

[7] Marjorie Taylor, "Conceptual Perspective Taking: Children's Ability to Distinguish What They Know from What They See," *Child Development* 59 (3) (June 1988): 703 – 718.

[8] Jennifer M. Jenkins and Janet Wilde Astington, "Cognitive Factors and Family Structure Associated with Theory of Mind Development in Young Children," *Developmental Psychology* 32 (1) (January 1996): 70 – 83.

[9] V. Slaughter and Alison Gopnik, "Conceptual Coherence in the Child's Theory of Mind," *Child Development* 67 (6) (1996): 2967 – 2989.

ries, an analogy I've called the theory theory. Like scientists, children seem to develop a succession of related intuitive causal theories of the world, theories that they expand, elaborate, modify, and revise in the light of new evidence.

There is only one problem with the theory theory, and it harks back to Popper's talk at Cambridge. We have had almost no idea how scientists learn about the world; when we 'theory theorists' turned to philosophers of science to find out about scientific learning mechanisms, we got the runaround. Philosophers knew that insofar as a theory was a deductive system, you could say something about how one part of the theory should follow from another; and they knew something, though much less, about how evidence could confirm or falsify a hypothesis that had been generated by a theory (this, of course, was where Popper made his contribution).

But they knew almost nothing about what has been called the logic of discovery – the way that experience itself might lead to the generation of new theories or hypotheses. And notoriously, they knew even less about what psychologists call conceptual changes (and what the rest of the world, ad nauseam, calls paradigm shifts), in which the very vocabulary of a theory seems to change in the light of new evidence. Some philosophers said that to answer questions about discovery and conceptual change you would have to go talk to psychologists. Others, even more discouragingly, said the questions were simply unanswerable. And if there were no accurate learning mechanisms that underlaid science, if Wittgenstein was right that the problems of induction, discovery, and conceptual change were not solvable, then the whole enterprise of science was in doubt.

So philosophers of science and developmental psychologists have been in the same unfortunate boat, convinced that the scientists and children they study are getting to the truth, perhaps even suspecting that they may be using some of the same learning mechanisms to get there, but unable to determine how. So both groups have mostly ended up waving their hands and talking vaguely about paradigm shifts and constructivism.

Ten years ago I would have said that this sad state of affairs was irremediable, at least for the immediate future. Our generation of scientists would have to labor over the details of the empirical natural history of learning and leave it to the next generation to develop precise and convincing explanations of learning. But, rather remarkably, age has made me more optimistic. Though we are still very far from having the whole story, I think there is a new line of work that is actually on the right track. We are beginning to understand not only what babies (and scientists) know when – but also *how* they learn it and *why* they get it right.

The general structure of the explanation comes from an entirely different part of cognitive science: the study of vision.[10] Indeed, the study of vision has been the most striking, though unheralded, success story in cognitive science – a case of real rather than just-so evolutionary psychology. Although we don't typically think of vision as a kind of learning, there is a sense in which the two processes are quite similar. The visual system takes a pattern of retinal input and generates accurate representations of three-dimensional objects moving through space. It has to solve what has been called the inverse problem: the three-dimensional world produces cer-

[10] Stephen E. Palmer, *Vision Science: Photons to Phenomenology* (Cambridge, Mass.: MIT Press, 1999).

tain patterns at the retina and the brain has to work backward to accurately recreate the world from that information. We have a remarkably good understanding of the computations, and even the neurological mechanisms, that are involved in this process.

The visual system solves the inverse problem by making certain very abstract and general assumptions about how the three-dimensional world creates patterns on the retina. And we can explain the way the system works by describing it in terms of these assumptions, and in terms of knowledge, rules, and inferences – just as we can explain how my computer works in this way. For example, the visual system seems to assume that the images at the retina of each eye are projections of the same three-dimensional objects in the world, and that the discrepancies between them are the result of geometry and optics. We can show mathematically that, given these assumptions, only some three-dimensional configurations of objects, and not others, will be compatible with a particular set of retinal patterns. This enables us to also say mathematically whether a visual system (human, animal, or robotic) generates the right representations of the spatial world from a particular pattern of data. In fact, the human visual system seems to be about as good at getting the right representations as it could possibly be.

The assumptions that allow these inferences to take place are themselves contingent and sometimes may be violated. For example, the View-Master toys and 3-D glasses of my youth and their modern virtual reality equivalents artificially create retinal images that normally would be generated by three-dimensional objects, and the visual system gets it wrong as a result. We see a three-dimensional Taj Mahal or oncoming train rather than two slightly different two-dimensional photographs.

But the consequences of those assumptions are deductive. It is not always true that retinal images are generated by light reflecting off the same three-dimensional object onto two separate retinas. But if it is true, then we can say, as a geometrical fact, that only certain kinds of images will result. In fact, of course, in real life, without the demonic View-Master to confuse things, the assumptions of the visual system will almost always be correct. That's why the designers of computer vision systems build those assumptions into their programs, and presumably that's why evolution built those assumptions into the design of the visual cortex.

In learning, as in vision, our brains may be performing computations that we can't perform consciously. We see a three-dimensional world or know about a causal one, without having to bother about the implicit computations that let us generate that world from the data. In vision science, we figure out which computations the brain performs by giving people particular patterns of retinal data and recording what they see. In the same way, we can give babies and young children patterns of statistical data and record what they learn.

When trained scientists do statistics, we make certain very general assumptions about what the underlying causal structure of the world is like, and how that structure leads to particular patterns of data. The data we consider are patterns of dependence and independence among variables. Just looking at a single dependency between two variables may not tell us a great deal about causal structure, just as looking at a small piece of a picture won't tell us much about a spatial scene. But by looking at the entire pattern of dependence and independence among several types of variables, we can zero in on the right causal structure, and eliminate incorrect hypotheses. Sometimes we can even use these patterns to add to the vocabulary of the theory. For instance, if we find otherwise unexplained dependencies between two variables, we may decide that there is a hidden unobserved variable that influences them both. Recently, philosophers of science, computer scientists, and statisticians working with what is called the Bayes net formalism have begun to provide a precise mathematical account of these kinds of inferences (see Clark Glymour's essay in this issue).

It turns out that even very young babies, as young as eight months old, are sensitive to patterns of dependency. We can play babies strings of syllables in various probabilistic combinations with particular patterns of dependency – for example, 'ba' may usually precede 'da,' but rarely precede 'ga.' The babies can use these patterns of probabilities to infer which combinations of syllables are likely to occur together, and they can also detect similar statistical patterns among musical tones or aspects of a visual scene. Babies also seem able to map those probabilities onto representations of the external world. They don't, for example, just notice that certain syllables tend to go together; they assume that these regularities occur because these combinations of syllables constitute words in the language they hear around them. In the example above, they would assume that 'bada' is more likely to be a word than 'baga.'[11]

We have shown that, at least by the time they are two and a half, children can also use patterns of conditional probability to make genuinely causal inferences. To do this, we show children a machine called the blicket detector. The machine is a square box that lights up and plays music when particular blocks are placed on top of it. The blocks are all different from one another, so the job for children is to identify which blocks are blickets, that is, which blocks will cause the machine to light up. We can present the children with quite complex patterns of contingency between the activation of the detector and various combinations of blocks. We can ask them which blocks are blickets, and we can ask them to activate the machine or get it to stop. And their answers are almost always correct. They make the right inferences about the causal powers of the blocks. They make the sort of statistical inferences a scientist would make and, according to the Bayes net formalism, should make. In similar experiments, we can even show that children postulate unobserved variables to deal with otherwise inexplicable patterns of data.[12]

In order to make inferences about the causal structure of the world and causal relations among variables, the scientist performs experiments. The scientist intentionally intervenes on a variable in the world, forcing it to have a particular value and then observing what happens to the values of other variables. Again Bayes nets provide a precise mathematical account of such inferences.

In a similar way, even the youngest babies are particularly sensitive to the consequences of their interventions on

[11] Richard N. Aslin, Jenny R. Saffran, and Elissa L. Newport, "Computation of Conditional Probability Statistics by 8-Month-Old Infants," *Psychological Science* 9 (4) (July 1998): 321–324.

[12] Alison Gopnik, Clark Glymour, David Sobel, Laura Schulz, Tamar Kushnir, and David Danks, "A Theory of Causal Learning in Children: Causal Maps and Bayes-Nets," *Psychological Review* 111 (1) (2004): 1–30.

the world. For example, with a ribbon we can attach a mobile to a three-month-old baby's leg; the baby will regard her influence over the mobile with fascination, systematically exploring the contingencies between various limb movements and the movements of the mobile.[13] By the time they are a year old, babies will systematically vary the kinds of actions they perform on objects, as they simultaneously observe the consequences of those actions. And they may watch the further consequences of the action 'downstream' and use that information to design new actions. Give a one-year-old a set of blocks and you can see her trying different combinations, placements, and angles, and gauging which of these will produce stable towers and which will end in equally satisfying crashes.

We have shown that by the time children are four they will intervene in the world in a way that lets them uncover causal structure. My student Laura Schulz's gear toy tests show how children learn about causal structure. This toy, like the blicket detector, presents children with a new causal relation that they must infer from evidence about contingencies. It is a square box with two gears on top and a switch on the side. When you flip the switch the gears turn simultaneously. If you remove gear A and then flip the switch, B turns by itself; if you remove gear B and flip the switch, A doesn't turn. With both of these pieces of evidence you can conclude that B is making A move. We tell the children that one of the gears makes the other one move, and then leave them alone with the toy and a hidden camera. The children swiftly produce the right set of experimental interventions with gear and switch to determine which gear moves the other.

Of course these observations will not surprise anyone who has spent much time with infants or young children, who are perpetually 'getting into things.' In this sense, we may think of toddlers as causal learning machines. They are small human versions of the Mars rovers that roam about getting into things on the red planet – except that children are also mission control, interpreting the data they collect.

Somewhere between statistical observation and active experimentation, scientists and babies alike learn from the interventions of others. Scientists read journals, go to talks, hold lab meetings, and visit other labs – and all those conferences surely have some function beyond assortative mating. We scientists make the assumption that the interventions of others are like our own interventions, and that we can learn similar things from both sources.

By at least nine months, human infants seem to make the same assumption. For example, in one study babies see an experimenter enter the room and touch the top of his head to a box that then lights up. A day later, babies return to the room, see the box, and then immediately touch their heads against the top of it.[14]

We have shown that by four, children can use information about the interventions of others appropriately to make new causal inferences. Consider the gear toy experiment described above. Children will also solve this task if they simply see an adult perform the right experiments on the toy. They not only learn

[13] Carolyn Rovee-Collier and Rachel Barr, "Infant Learning and Memory," in Gavin Bremner and Alan Fogel, eds., *Blackwell Handbook of Infant Development* (Malden, Mass.: Blackwell Publishers, 2001), 139 – 168.

[14] Andrew N. Meltzoff, "Infant Imitation and Memory: Nine-Month-Olds in Immediate and Deferred Tests," *Child Development* 59 (1) (1988): 217 – 225.

about the causal consequences of adult actions, but also about the causal relations among the objects upon which adults perform those actions.

Indeed, the three techniques of causal inference that I have described – analyzing statistics, performing experiments, and watching the experiments of others – may give both scientists and children their extraordinary learning powers. Elements of the first two techniques are probably in place even in nonhuman animals. In classical conditioning, animals calculate dependencies among particularly important events, like shock and food. In operant conditioning, animals calculate the consequences of their actions. This is not surprising given the importance of causal knowledge for survival.

However, as Mike Tomasello and Danny Povinelli point out in this issue, there is much less clear evidence of the third type of learning – learning from the actions of others – in other animals. And there is no evidence that other animals combine all three types and assume that they provide information about the causal structure of the external world. By contrast, human children, at least by age three or four, do seem to put these types of information together in this way. This ability may, in fact, be one of the crucial abilities that give human beings their unique intellectual capacities. It allows them to learn far more about the world around them than other animals, and to use that knowledge to change the world.

My guess is that many of the mistakes that children and adults make in learning don't happen because they make the wrong deductions from assumptions and evidence, but rather because they make assumptions that are unwarranted under the particular circumstances.

For example, children tend to assume that the samples of evidence they collect are representative of the data. Similarly, they seem to assume that their own actions and the actions of others have all the formal characteristics of an ideal experimental intervention. The self-conscious methodological canons of formal science – the courses on statistics and experimental design – are intended to make these assumptions explicit rather than implicit and so ensure that they are correct in particular cases. For children, however, the assumptions may be close enough to the truth most of the time, and the evidence may be sufficiently rich, so that they mostly get things right anyway.

If we want children, and lay adults, to understand and appreciate science, we may need to make more connections between their intuitive and implicit causal inference methods and the self-conscious and explicit use of these methods in science. We may need, literally, a sort of scientific consciousness-raising.

Popper's quarrel with Wittgenstein reflected a larger argument between the view that science and philosophy tell us new things about the world, and the view that all they do is reflect social arrangements and linguistic conventions.

If we could put children in touch with their inner scientists, we might be able to bridge the divide between everyday knowledge and the apparently intimidating and elite apparatus of formal science. We might be able to convince them that there is a deep link between the realism of everyday life and scientific realism. And if we were able to do that, then we might win Popper's argument for him – without having to resort to pokers.

Daniel John Povinelli

Behind the ape's appearance: escaping anthropocentrism in the study of other minds

Look at Megan. Not just at her distinctively chimpanzee features – her accentuated brow ridge, her prognathic face, her coarse black hair – but at the totality of her being: her darting eyes, her slow, studied movements, the gestures she makes as her companion, Jadine, passes nearby. Can there be any doubt that behind certain obvious differences in her appearance resides a mind nearly identical to our own? Indeed, is it even possible to spend an afternoon with her and *not* come to this conclusion? Upon reflection, you will probably acknowledge that her mind is not identical to ours. "But surely it's not qualitatively different, either," you will still insist. "I mean, it's obvious from watching her that we share the same *kind* of mind."

Faced with the overwhelming similarity in the spontaneous, everyday behavior of humans and chimpanzees, how can someone like me – someone who has dedicated his life to studying these remarkable animals – entertain the possibility that their minds are, in profound respects, radically different from our own? How can I challenge the received wisdom of Darwin – confirmed by my own initial impressions – that the mental life of a chimpanzee is best compared to that of a human child?

Actually, it's easy: I have learned to have more respect for them than that. I have come to see that we distort their true nature by conceiving of their minds as smaller, duller, less talkative versions of our own. Casting aside these insidious assumptions has been difficult, but it has allowed me to see more clearly that the human mind is not the gold standard against which other minds must be judged. For me it has also illuminated the possibility of creating a science that is less contaminated by our deeply anthropocentric intuitions about the nature of other minds.

The best available estimates suggest that humans and chimpanzees originated from a common ancestor about five or six million years ago.[1] This is reflected

Daniel John Povinelli is Louisiana Board of Regents Endowed Professor of Science at the University of Louisiana at Lafayette, and director of the Cognitive Evolution Group and the Center for Child Studies. His latest book is "Folk Physics for Apes: The Chimpanzee's Theory of How the World Works" (2000).

[1] Galina V. Galzko and Masatoshi Nei, "Estimation of Divergence Times for Major Lineages of Primate Species," *Molecular Biology and Evolution* 20 (2003): 424 – 434.

in estimates of our genetic similarity: we share, on average, about 98.6 percent of our total nucleotide sequence in common. This statistic seems impressive. After all, such biological affinity would appear to be the final nail in the coffin of the notion that there could be any radical mental differences between them and us: if chimpanzees and humans share 98.6 percent of their genetic material, then doesn't it follow that there ought to be an extraordinarily high degree of mental similarity as well? This idea has been paraded so frequently through the introductory paragraphs of both scholarly journal articles and the popular press alike that it has come to constitute a melody of sorts; an anthem that if not sung raises doubts as to one's allegiance to the cause of defending the chimpanzee's dignity.

But what does this 98.6 percent statistic really mean? It should be of immediate interest that it is almost invariably misreported. We do not share 98.6 percent of our genes in common with chimpanzees; we share 98.6 percent of our nucleotide sequence. A single nucleotide difference in a string of four hundred may code for a different allele. Furthermore, as the geneticist Jonathan Marks has pointed out in lucid detail, the 98.6 percent statistic has so little grounding in the average mind that confronts it, as to render it essentially meaningless.[2] We might, after all, share 50 percent of our nucleotide sequences in common with bananas and broccoli. But what on earth does it mean to say that we are 50 percent the same as a vegetable? I don't know about you, but I doubt my mind is 50 percent identical to that of the garden pea. And so what would it mean, exactly, if we discovered that our minds were 75 percent chimpanzee?

No, such coarse genetic comparisons will hardly suffice to help us understand the complex similarities and differences that exist between the mental lives of humans and chimpanzees. However, in a climate where certain highly visible experts have radically anthropomorphized chimpanzees,[3] such statistics are heralded as establishing once and for all that chimpanzees are, at the very least, mentally equivalent to two- or three-year-old human children, and should therefore be granted human rights.[4]

A few obvious biological facts may be worth noting here. To begin, it was the human lineage, not the chimpanzee one, that underwent radical changes after our respective geneologies began to diverge from their common ancestor. Since this split, humans have resculpted their bodies from head to toe – quite literally, in fact; as our lineage became bipedal, the pelvis, the knee, and the foot were all drastically reshaped, with modifications in the hand (including new muscles) soon following. To top it all off, we ultimately tripled the size of our brain, with disproportionate increases probably occurring in the seat of higher cognitive function, the prefrontal cortex. Oh yes, and at some point during all of this (no one knows exactly when), natural language – perhaps the most notice-

[2] Jonathan Marks, *What It Means to Be 98% Chimpanzee* (Berkeley: University of California Press, 2002).

[3] For examples, see Sue Savage-Rumbaugh, *Kanzi: The Ape at the Brink of the Human Mind* (New York: John Wiley & Sons, 1994); Jane Goodall, *Through a Window* (Boston: Houghton Mifflin, 1990); Roger Fouts, *Next of Kin* (New York: William Morrow and Co., 1997).

[4] Steven M. Wise, *Rattling the Cage: Toward Legal Rights for Animals* (Cambridge, Mass.: Perseus Books, 2000); Paola Cavalieri and Peter Singer, eds., *The Great Ape Project: Equality Beyond Humanity* (New York: St. Martin's Press, 1993).

able of human adaptations – emerged as well.

In contrast, chimpanzees have probably changed relatively little from the common ancestor they shared with us about five million years ago. Indeed, of all of the members of the great ape/human group who shared a common ancestor about fifteen million years ago, none, indeed, has diverged as much as humans. A simple thought experiment may help to put this point into perspective: line up all of the species in question – gorillas, orangutans, chimpanzees, bonobos, humans – and one of them immediately stands out. Guess which one?

In fact, the more we compare humans and chimpanzees, the more the differences are becoming apparent. Even geneticists are starting to catch up with the reality of these differences. New research has shown that rough similarity in our nucleotide sequences obscures the fact that the same genes may have dramatically different activity levels in the two species. So even where humans and chimpanzees share genes in common, it turns out that there are what can only be described as major differences in gene *expression* – that is, whether, when, and for how long genes are actually working to produce the proteins for which they code.[5] This is the *real* stuff of genetic comparison, and it casts our crude genetic similarity to the garden pea in a wholly different light.

What makes these differences in gene expression significant is that they ultimately manifest themselves as differences in the bodies – including the brains – of humans and chimpanzees. So, exactly how similar are the brains of humans and chimpanzees? After all, if we knew *that*, couldn't we directly address the question of their mental similarity? Well, it would be a start, anyhow. Unfortunately, comparisons of the brains of humans and apes have traditionally been limited to gross considerations such as size and surface features (such as lobes and sulcus patterns). Remarkably, the details of the internal organization of human and great ape brain systems and structures have been largely ignored, in part because it's so difficult to study these brains, but also because most neuroscientists have frequently assumed that despite great differences in size, all mammalian brains are organized pretty much the same.

Fortunately, even this is beginning to change. For example, Todd Preuss, working at the University of Louisiana, recently made a startling discovery while comparing the brains of humans and chimpanzees. Turning his attention away from the frontal lobes, his previous area of research, Preuss decided to take a look at the primary visual cortex (V1), the area of the cerebral cortex that is the first way station into the processing of visual information. The organization of this area of the brain has been assumed to be nearly identical across primates. But there, in the middle of V1, Preuss and his colleagues uncovered a distinctively human specialization – a kind of neural architecture not found even in chimpanzees.[6] Preuss speculates that this specialization involves modifications of the pathways related to spatial vision and motion processing. But, regardless of what it is for, it suggests that

[5] Wolfgang Enard et al., "Intra- and Interspecific Variation in Primate Gene Expression Patterns," *Science* 296 (2002): 341 – 343; Mario Cáceres et al., "Elevated Gene Expression Levels Distinguish Human from Non-Human Primate Brains," *Proceedings of the National Academy of Sciences* 100 (2003): 13030 – 13035.

[6] Todd M. Preuss et al., "Distinctive Compartmental Organization of Human Primary Visual Cortex," *Proceedings of the National Academy of Sciences* 96 (1999): 11601 – 11606.

we need to rethink brain evolution in a way that's consistent with neo-Darwinian theory: similarity and difference among species as comfortable bedfellows; a state of affairs accomplished by weaving in new systems and structures alongside the old. "If we find such differences in the middle of the primary visual cortex," Preuss recently remarked to me, "just imagine what we're going to find when we start looking elsewhere."

Some may be surprised (or even afraid) to learn of such differences between humans and our nearest living relatives. After several decades of being fed a diet heavy on exaggerated claims of the degree of mental continuity between humans and apes, many scientists and laypersons alike now find it difficult to confront the existence of radical differences. But then, in retrospect, how viable was the idea of seamless mental continuity in the first place? After all, it tended to portray chimpanzees as watered-down humans, not-quite-finished children. Despite the fact that aspects of this notion can be traced straight to Darwin, it is an evolutionarily dubious proposition, to say the least.

If there are substantial differences between the mental abilities of humans and chimpanzees, in what areas are they likely to exist? Over the past couple of thousand years, many potential rubicons separating human and animal thinking have been proposed. Some of these have been particularly unhelpful, such as the radical behaviorists' forgettable proposition that animals don't 'think' at all (of course, these behaviorists were even skeptical about the existence of human thought!). And, unfortunately, in the popular imagination the question still appears to be, "*Can* animals think?"[7] as opposed to, "How does thinking differ across species?" (the latter being a decidedly more evolutionarily minded question).

Assuming that chimpanzees and other species have mental states (a point I take for granted), it seems to me that a more productive question to ask is, "What are their mental states about?" Or, put another way, "What kinds of concepts do they have at their disposal?" It would stand to reason that the mental states of chimpanzees, first and foremost, must be concerned with the things most relevant to their natural ecology – remembering the location of fruit trees, keeping an eye out for predators, and keeping track of the alpha male, for instance. And so surely chimpanzees form concepts about concrete things – things like trees, facial expressions, threat vocalizations, leopards, and the like. But what about more abstract concepts? Concepts like ghosts, gravity, and God?

Admittedly, to use the term 'concept' as loosely as I have will require the indulgence of certain scholars. But perhaps some progress can be made by noting that every concept is at least somewhat abstract if it extends beyond a particular example. For instance, if one has a notion of an apple that is not limited to a single instance of that apple, then one has made a generalization, and thus a kind of abstraction. Given that it has been known for decades or more that chimpanzees and many other species form such abstractions,[8] this cannot be a defining feature of human thinking.

[7] Eugene Linden, "Can Animals Think?" *Time*, 22 March 1993.

[8] Suzette L. Astley and Edward A. Wasserman, "Object Concepts: Behavioral Research with Animals and Young Children," in William T. O'Donohue, ed., *Learning and Behavior Therapy* (Boston: Allyn and Bacon, 1997), 440 – 463; Tom R. Zentall, "The Case for a Cognitive Approach to Animal Learning and Behavior," *Behavioural Processes* 54 (2001): 65 – 78.

Behind the ape's appearance

At the risk of oversimplification, let me instead propose a distinction between concepts that refer to objects and events that can be directly observed (that is, things that can be detected by the unaided senses), versus hypothetical entities and processes (things that are classically unobservable). Thus, I wish to separately consider all concepts that refer to theoretical things: all the things that are not directly registered by the senses, but are merely posited to exist on the basis of things we can observe.

Such concepts permeate our commonsense way of thinking: we explain physical events on the basis of things like 'forces' (supernatural or otherwise) that we have never actually witnessed, and account for the behavior of other humans on the basis of mental states we have never seen (e.g., their beliefs, desires, and emotions). These concepts serve as the bedrock for some of our most fundamental explanations for why the world works the way it does.

Meanwhile, we can directly contrast these sorts of concepts with ones that are derived from things that can be directly observed: apples and oranges, trees, flashes of lightning, facial expressions – even the raising of a hand or the sound of a train whistle blowing in the distance. Concepts about these things share at least one property in common: they are all derived from the world of macroscopic entities with which the primary senses directly interact. Without additional justification, I am therefore asserting a distinction between concepts that refer to observable objects and events, and ones that refer to strictly hypothetical ones.

So, here's a proposal: the mental lives of humans and chimpanzees are *similar*, in that both species form innumerable (and in many cases, identical) concepts about observable things; but, at the same time, are *radically different*, in that humans form additional concepts about inherently unobservable things.[9]

Now, I realize that most people would not be surprised if it were established beyond doubt that chimpanzees lack a concept of God. But what about other, seemingly more prosaic concepts that infest our way of thinking about the world? Consider the way in which we think about the social realm. In interacting with each other (and with animals, for that matter), we use a dual system of representation: we understand other beings both as part of the observable world (they engage in particular movements of their hands and feet, and their lips form particular contortions as sounds emerge from their mouths), and as entities with mental properties – unobservable attributes like emotions, intentions, desires, and beliefs.

The proposal is that, in contrast to humans, chimpanzees rely strictly upon observable features of others to forge their social concepts. If correct, it would mean that chimpanzees do not realize that there is more to others than their movements, facial expressions, and habits of behavior. They would not understand that other beings are repositories of private, internal experience. They would not appreciate that in addition to things that go on in the observable world, there are forever hidden things that go on in the private life of the mind. It would mean that chimpanzees do not reason about what others think, believe, and feel – precisely because they do not form such concepts in the first place.

[9] This discussion extends several previous descriptions of this hypothesis, for example, my article with Jesse Bering and Steve Giambrone, "Toward a Science of Other Minds: Escaping the Argument by Analogy," *Cognitive Science* 24 (2000): 509 – 541.

Before we get too much further, let me be honest: I recognize that this proposal has troubling implications. For one thing, if chimpanzees do not reason about unobservable entities, then we would frequently need distinctly different explanations for human and chimpanzee behavior – even in situations where the behavior looks almost identical. Mind you, we would not need completely different explanations, just ones that are distinctive enough to capture the proposed difference. Nonetheless, each time we witnessed a chimpanzee engage in a complex social behavior that resembles our own, we would have to believe that, unlike us, the chimpanzee has only one conceptual system for encoding and reasoning about what is happening: a system that invokes concepts derived from observable features of the world. Thus, when chimpanzees deceive each other (which they do regularly), they would never be trying to manipulate what others believe, nor what others can see or hear, for constructs like 'believing,' 'seeing,' and 'hearing' are already deeply psychological. No, in deciding what to do, the chimpanzee would be thinking and reasoning solely about the abstracted statistical regularities that exist among certain events and the behaviors, postures, and head movements (for example) of others – what we have called 'behavioral abstractions.'[10]

I should note that humans, too, rely heavily upon behavioral abstractions in their day-to-day interactions. We *must* be doing so: otherwise upon what basis could we attribute additional, psychological states to others? First, we recognize the turn of the head and the direction of the eyes (observable features), then we ascribe the internal experience of 'seeing' (unobservable feature). So, the proposal isn't that chimpanzees use one system and humans use another; both species are purported to rely upon concepts about the observable properties of others. Instead, the proposal is that chimpanzees don't form additional concepts about the *un*observable properties of other beings (or the world in general, for that matter).

So, at face value, the proposal I have made is worrying. In interpreting what would appear to be the exact same behaviors in humans and chimpanzees in different ways, I seem to be applying a double standard.

But is this implication really problematic, or does it just seem problematic because it runs counter to some of our most deeply engrained – but fundamentally flawed – ways of thinking?

Assume, for a moment, that you have traveled back in time to a point when there were no chimpanzees on this planet – and no humans, either. Imagine further that you have come face to face with members of the last common ancestor of humans and chimpanzees. Let's stipulate that these organisms are intelligent, thinking creatures who deftly attend to and learn about the regularities that unfold in the world around them. But let us also stipulate that they do not reason about unobservable things; they have no ideas about the 'mind,' no notion of 'causation.'

As you return to your time machine and speed forward, you will observe new lineages spring to life from this common ancestor. Numerous ape-like species will emerge, then disappear. As you approach the present day, you will even witness the evolutionary birth of modern orangutans, chimpanzees, and gorillas. But amid all of this your attention will be drawn to one particular offshoot of this

[10] Daniel J. Povinelli and Jennifer Vonk, "Chimpanzee Minds: Suspiciously Human?" *Trends in Cognitive Science* 7 (2003): 157 – 160.

process, a peculiar genealogy that buds off numerous descendent species. This particular lineage has evolved an eye-catching trick: it habitually stands upright; it walks bipedally. And some of its descendants build upon this trick, capitalizing upon the new opportunities it offers. For reasons that we may never fully know, tool use and manufacture increase exponentially, language emerges, brain size triples, and, as more time passes, human material and social culture begins to accrete upon the shoulders of the lineage's last surviving member: *Homo sapiens sapiens*. Now, imagine that as part of this process, this lineage evolved new conceptual structures (intimately connected to the evolution of language) that allow them to reason about things that cannot be observed: mental states, physical forces, spiritual deities.

I have stipulated all of this so we can confront the following question: If evolution proceeded in this quite plausible manner, then how would we expect the spontaneous, everyday behavior of humans to compare to that of chimpanzees? The answer, I think, is that things would look pretty much the way they do now. After all, humans would not have abandoned the important, ancestral psychological structures for keeping track of other individuals within their groups, nor jettisoned their systems for noticing that something very different happens when Joe turns his head toward so-and-so, just depending on whether or not his hair is standing on end. No, in evolving a new psychological system for reasoning about hypothetical, internal mental states, humans would not have (indeed, could not have!) abandoned the ancient systems for reasoning about observable behavior. The new system by definition would depend upon the presence of older ones.

Now, is it really troubling to invoke a different explanation for what on the surface seem to be identical units of behavior in humans and chimpanzees? If the scenario I have outlined above is correct, then the answer must be, no. After all, for any given ability that humans and chimpanzees share in common, the two species would share a common set of psychological structures, which, at the same time, humans would augment by relying upon a system or systems unique to our species. The residual effect of this would manifest itself in numerous ways: some subtle (such as tightly constrained changes in the details of things to which our visual systems attend), others more profound (such as the creation of cultural artifacts like the issue of *Dædalus* in which you are now reading these words).

Behind the ape's appearance

So much for theory. What about the empirical evidence; does it support the proposal I have just offered? Although it will not surprise you to learn that I think it does, I have not always been of this opinion; I used to believe that any differences between humans and chimpanzees would have to be trivial. But the results of over two hundred studies that we have conducted during the past fifteen years have slowly changed my mind. Combined with findings from other laboratories, this evidence has forced me to seriously confront the possibility that chimpanzees do not reason about inherently unobservable phenomena.

Let me briefly illustrate this evidence with three simple examples: one from the social domain, one from the domain of physics, and one from the domain of numerical reasoning.

First, what does the experimental evidence suggest about whether chimpanzees reason about mental states? Although the opinions of experts differ (and have swung back and forth over the past several years), I believe that at present there is no direct evidence that chimpanzees conceive of mental states, and considerable evidence that they do not. As an example, consider the well-studied question of whether chimpanzees reason about the internal, visual experiences of others, that is, of whether they know anything about 'seeing.'

To begin, no one doubts that chimpanzees respond to, reason about, and form concepts related to the movements of the head, face, and eyes of others; these are aspects of behavior that can be readily witnessed.[11] But what about the idea that another being 'sees' things, that others are loci of unobservable, visual experiences?

Over the past ten years we have conducted dozens of studies of juvenile, adolescent, and adult chimpanzees to explore this question. Perhaps the most straightforward of these studies involved examining how chimpanzees understand circumstances under which others obviously can or cannot see them.[12] In these studies, chimpanzees were exposed to a routine in which they would approach a familiar playmate or caretaker to request a food treat using their species-typical begging gesture. Simple enough. But on the crucial test trials, the chimpanzees were confronted with two individuals, only one of whom could see them. For example, in one condition, one caretaker had a blindfold covering her mouth, whereas the other had a blindfold covering her eyes. The question was to whom would the chimpanzee gesture.

Not surprisingly, in our trials with human children, even two-year-olds gestured to whoever had the blindfold over her mouth (versus the eyes), probably because they could represent her inner, psychological state ("She can *see* me!"). In striking contrast, our chimpanzees did nothing of the kind. Indeed, in numerous studies, our chimpanzees gave virtually no indication that they could understand 'seeing' as an internal experience of others.

With enough trials of any given condition the chimpanzees were able to *learn* to select whoever was able to see them; after enough trials of not being handed a banana when gesturing to someone with a bucket over her head, the chimpanzees figured out to gesture to the other person. Did this mean that they had finally discerned what we were asking them? In numerous transfer tests in which we pitted the idea that the chimpanzees were learning about the observable cues (i.e., frontal posture, presence of the face or eyes) against the possibility that on the basis of such cues they were reasoning about who could 'see' them, the chimpanzees consistently insisted (through their behavior) that they were reasoning about observable features, not internal mental states, to guide their choices.

In addition to what they learned in these tests, it also became apparent that chimpanzees come pre-prepared, as it

[11] See Daniel J. Povinelli and Timothy J. Eddy, "Chimpanzees: Joint Visual Attention," *Psychological Science* 7 (1996): 129 – 135; Shoji Itakura, "An Exploratory Study of Gaze-Monitoring in Nonhuman Primates," *Japanese Psychological Research* 38 (1996): 174 – 180; Michael Tomasello, Brian Hare, and Josep Call, "Five Primate Species Follow the Visual Gaze of Conspecifics," *Animal Behaviour* 58 (1998): 769 – 777.

[12] Our laboratory's empirical research of chimpanzees' understanding of 'seeing' has been summarized in my article, "The Minds of Humans and Apes are Different Outcomes of an Evolutionary Experiment," in Susan M. Fitzpatrick and John T. Bruer, eds., *Carving Our Destiny: Scientific Research Faces a New Millennium* (Washington, D.C.: National Academy of Sciences and John Henry Press, 2001), 1 – 40.

were, to make sense of certain postures. For instance, in our tests they immediately knew what to do when confronted with someone facing them versus someone facing away, and this finding has been replicated in several other laboratories.[13] "But if they make *that* distinction," you wonder, "then why do they perform so differently on the other tests? Is it just because they're confused? How are we to make sense of such a puzzling pattern of findings?"

Actually, these results are not puzzling at all if the ability to reason about mental states evolved in the manner that I suggested earlier – that is, if humans wove a system for reasoning about mental states into an existing system for reasoning about behavior. After all, if the idea is correct, then chimpanzees may well be born predisposed to attend to certain postures and behaviors related to 'seeing' – even though they know nothing at all about such mental states *per se* – precisely because overt features of behavior are the tell-tale indicators of the future behavior of others. But when such features are carefully teased apart to probe for the presence of a mentalistic construal of others, the chimpanzees stare back blankly: this is not part of their biological endowment. Thus, if the evolutionary framework I have sketched is correct, neither the chimpanzees nor the results are 'confused'; that epithet may fall squarely upon the shoulders of we human experimenters and theorists who are so blinded by our own way of understanding the world that we are not readily open to the chimpanzee's way of viewing things.

Of course, some have challenged this conclusion, arguing that we need to turn up the microscope and develop more tests that will allow chimpanzees to express their less well-developed understanding of such concepts.[14] So, for example, researchers at Emory University recently conducted tests in which a dominant and a subordinate chimpanzee were allowed to fight over food that was positioned in an enclosure between them.[15] On the critical trials, two pieces of food were positioned equidistant from the animals. The catch was that one piece of food was placed behind an opaque barrier so that only the subordinate could see it. The researchers report that when the subordinate was released into the enclosure, he or she tended to head for the food that was hidden from the dominant's view, suggesting, perhaps, that the subordinate was modeling the visual experience of his or her dominant rival.

Behind the ape's appearance

But do such tests really help?[16] Do they reveal some weaker understanding

[13] For example, see Autumn B. Hostetter et al., "Differential Use of Vocal and Gestural Communication by Chimpanzees (*Pan troglodytes*) in Response to the Attentional Status of a Human (*Homo sapiens*)," *Journal of Comparative Psychology* 115 (2001): 337 – 343.

[14] Michael Tomasello et al., "Chimpanzees Understand Psychological States – The Question Is Which Ones and to What Extent," *Trends in Cognitive Science* 7 (2003): 153 – 156, esp. 156.

[15] Brian Hare et al., "Chimpanzees Know What Conspecifics Do and Do Not See," *Animal Behaviour* 59 (2000): 771 – 785; see also M. Rosalyn Karin-D'Arcy and Daniel J. Povinelli, "Do Chimpanzees Know What Each Other See? A Closer Look," *International Journal of Comparative Psychology* 15 (2002): 21 – 54.

[16] In a recent analysis of the diagnostic potential of these and other tests, Jennifer Vonk and I (see footnote 10) argued that the logic of current tests with chimpanzees (and other animals) cannot, in principle, provide evidence that uniquely supports the notion that they are reasoning about mental states (as opposed to behavior alone), and we advocated a new paradigm of tests that may have such diagnostic power. An alternative point of view is provided in the companion piece by Tomasello and col-

of mental states in chimpanzees? These are precisely the situations in which chimpanzees will be evolutionarily primed to use their abilities to form concepts about the actions of others to guide their social behavior. So, for example, they can simply know to avoid food that is out in the open when a dominant animal is about to be released. "But still," the skeptic within you asks, "that's pretty smart, isn't it? The chimpanzees would have to be paying attention to who's behind the door, and what that other individual is going to do when the door opens, right?"

Fair enough. But that, in the end, is the point: chimpanzees can be intelligent, thinking creatures even if they do not possess a system for reasoning about psychological states like 'seeing.' If it turns out that this is a uniquely human system, this should not detract from our sense of the evolved intelligence of apes. By way of analogy, the fact that bats echolocate but humans don't, hardly constitutes an intellectual or evolutionary crisis.

In the final analysis, the best theory will be the one that explains both data sets: the fact that chimpanzees reason about all the observable features of others that are associated with 'seeing' – and yet at the same time exhibit a striking lack of knowledge when those features are juxtaposed in a manner that they have never witnessed before (i.e., blindfolds over eyes versus over the mouth). I submit that, at least for the time being, the evolutionary hypothesis I have described best meets this criterion.

Table 1
Theoretical causal constructs and their observable 'ambassadors'

Theoretical concept	Paired observable 'ambassador'
gravity	*downward object trajectories*
transfer of force	*motion-contact-motion sequences*
strength	*propensity for deformation*
shape	*perceptual form*
physical connection	*degree of contact*
weight	*muscle/tendon stretch sensations*

A second example of the operation of what may be a uniquely human capacity to reason about unobservables comes from comparisons of humans' and chimpanzees' commonsense understanding of physics. Humans – even very young children – seem disposed to assume that there's more to the physical world than what meets the eye. For example, when one ball collides with another, stationary one, and the second speeds away, even quite young children are insistent that the first one caused the second to move away. Indeed, as Michotte's classic experiments revealed, this seems to be an automatic mental process in adult humans.[17] But what is it, exactly, that humans believe *causes* the movement of the second ball? As Hume noted long ago, they do not merely recognize that the objects touched; that's just a re-description of the observed events.[18] No, the first one is seen as hav-

leagues. However, I believe that this view dramatically underestimates the representational power of a psychological system that forms concepts solely about the observable aspects of behavior.

[17] Albert Michotte, *The Perception of Causality* (New York: Basic Books, 1963).

[18] David Hume, *Treatise of Human Nature*, vols. 1–2, ed. A. D. Lindsay (London: Dent, 1739; 1911).

ing transmitted something to the second object, some kind of 'force.' But where is this force? Can it be seen? No, it is a theoretical thing.

In an initial five-year study of 'chimpanzee physics,' we focused our apes' attention on simple tool-using problems.[19] Given their natural expertise with tools, our goal was to teach them how to solve simple problems – tasks involving pulling, pushing, poking, etc. – and then to use carefully designed transfer tests to assess their understanding of why the tool objects produced the effects they did. In this way, we attempted to determine if they reason about things like gravity, transfer of force, weight, and physical connection, or merely form concepts about spatio-temporal regularities. To do so, we contrasted such concepts with their perceptual 'ambassadors' (see table 1), much in the same way that we had contrasted the unobservable psychological state of 'seeing' against the observable behavioral regularities that co-vary with 'seeing.'

To pick just one example: we explored in detail the chimpanzee's understanding of physical connection – of the idea that two objects are bound together through some unobservable interaction such as the force transmitted by the mass of one object resting on another, or the frictional forces of one object against another; or conversely, the idea that simply because two objects are physically touching does not mean there is any real form of 'connection.' We presented our chimpanzees with numerous problems, but consider one test in which we first taught them to use a simple tool to hook a ring in order to drag a platform with a food treat on it toward them. Although they learned to do so, our real question was whether, when confronted with two new options, they would select the one involving genuine physical connection as opposed to mere 'contact.' Consistent with our findings in other tests, they did not. Instead, 'perceptual contact' seemed to be their operating concept. The observable property of contact (of any type) was generally sufficient for them to think that a tool could move another object.

Finally, consider the chimpanzee's numerical understanding. Over the past decade or so, it has become apparent that many species share what Stanislas Dehaene has called a 'number sense' – the ability to distinguish between larger and smaller quantities, even when the quantities being compared occupy identical volumes.[20]

In an attempt to explore the question of numerical reasoning in animals, several research laboratories have trained apes to match a specific quantity of items (say, three jelly beans) with the appropriate Arabic numeral.[21] That they can accomplish this should not be the least bit surprising: humans and chimpanzees (and many other species) share the ability to visually individuate objects. After extensive training, furthermore, the most apt of these pupils have gone on to exhibit some understanding of *ordinality* (the idea that 5 represents a

[19] Daniel J. Povinelli, *Folk Physics for Apes* (Oxford: Oxford University Press, 2000).

[20] Stanislas Dehaene, *The Number Sense* (Oxford: Oxford University Press, 1997).

[21] For this discussion, I rely heavily on the detailed results from Ai, a twenty-five-year-old chimpanzee whose numerical abilities have been studied since she was five by a team led by Tetsuro Matsuzawa in Kyoto, Japan. See Dora Biro and Tetsuro Matsuzawa, "Chimpanzee Numerical Competence: Cardinal and Ordinal Skills," in Tetsuro Matsuzawa, ed., *Primate Origins of Human Cognition and Behavior* (Tokyo: Springer, 2001), 199 – 225.

larger quantity than 4, for example). So, isn't this evidence that chimpanzees have a solid grasp of the notion of the number?

Let us scratch the surface a bit, to look at these findings from the perspective I have been advocating. First, do these chimpanzees possess a dual understanding of numbers – both as associates of real object sets and as inherently theoretical things – such that every successive number in the system is exactly '1' more than the previous number? The training data even from Ai, the most mathematically educated of all chimpanzees, suggests that they do not. For example, each time the next numeral was added into her training set, it took her just as long to learn its association with the appropriate number of objects as it took with the previous numeral. In other words, there appeared to be little evidence that Ai understood the symbols as anything other than associates of the object sets. Furthermore, even her dedicated mentors suggest that she was not 'counting' at all: with quantities of up to three or four objects, she performed like humans, using an automatic process ('subitizing') to make her judgments; but with larger quantities, instead of counting, it appears as if she was simply estimating 'larger' or 'smaller.'

What about ordinality? When first tested for her understanding of the relative ordering of numbers, Ai exhibited no evidence that this was part of her conceptual structure. That is, when presented with pairs of numbers, 1 versus 8, for example, she did not seem to have any notion that the value of 1 is smaller than the value of 8 – even though she had been correctly matching these numerals to object sets for years! Of course, after extended training, Ai did eventually exhibit evidence of this ability, and now, after more than fifteen years of training, when confronted with a scrambled array of the numerals 1 to 9, she has the remarkable ability to select them in ascending order.

But what does it mean that under the right training regime we can guide a chimpanzee like Ai into a performance that looks, in many but not all respects, like human counting? One possibility is that a basic number sense – a system grounded to individual macroscopic objects – is widespread among animals, and that apes (and other animals) can use this ability (in concert with their other cognitive skills) to figure out ways to cope with the 'rules' that humans establish in their tests. In contrast, the human system for counting (as well as other mathematical ideas) could be seen as building upon these older systems by reifying numbers as things in their own right – theoretical things. This may seem like a subtle and unimportant distinction for some tasks, but it may be one that leaves the ape mystified when facing questions that treat numbers as things in their own right.

As a striking example of the distinction I have been trying to draw, consider zero, surely one of the purest examples that exists of an inherently unobservable entity. If I am right, then zero ought to be virtually undetectable by the chimpanzee's cognitive system. And indeed, the data seem to bear this out.[22] For all of her training, even Ai does not appear to have learned to understand zero in this sense. True, she (and other animals) have quickly learned to pick the numeral 0 in response to the absence of objects (something easily explained by associative learning processes). But tests of or-

[22] Dora Biro and Tetsuro Matsuzawa, "Use of Numerical Symbols by the Chimpanzee (*Pan troglodytes*): Cardinal, Ordinals, and the Introduction of Zero," *Animal Cognition* 4 (2001): 193 – 199.

dinality involving zero (choosing whether 0 is greater or lesser than 6, for example) have consistently revealed what I believe might be best described as the virtual absence of the concept. Although this training has gradually forced her 'understanding' of zero into a position further and further down the 'number line,' even to this day, after thousands of trials, Ai still reliably confuses 0 with 1 (and in some tasks, with 2 or 3 as well). However one wishes to interpret such findings, they are certainly not consistent with an understanding of the very essence of zero-ness.[23]

Our work together is done. To the best of my ability I have laid out the case for believing that chimpanzees can be bright, alert, intelligent, fully cognitive creatures, and yet still have minds of their own. From this perspective, it may be our species that is the peculiar one – unsatisfied in merely knowing *what* things happen, but continually driven to explain *why* they happen, as well. Armed with a natural language that makes referring to abstract things easy, we continually pry behind appearances, probing ever deeper into the causal structure of things. Indeed, some tests we have conducted suggest that chimpanzees may not seek 'explanations' at all.[24]

And yet I cannot help but suspect that many of you will react to what I have said with a feeling of dismay – perhaps loss; a sense that if the possibility I have sketched here turns out to be correct, then our world will be an even lonelier place than it was before. But for the time being, at least, I ask you to stay this thought. After all, would it really be so disappointing if our first, uncontaminated glimpse into the mind of another species revealed a world strikingly different from our own; or all that surprising if the price of admission into that world were that we check some of our most familiar ways of thinking at the door? No, to me, the idea that there may be profound psychological differences between humans and chimpanzees no longer seems unsettling. On the contrary, it's the sort of possibility that has, on at least some occasions, emboldened our species to reach out and discover new worlds with open minds and hearts.

Behind the ape's appearance

[23] One might retort that the numeral 0 appeared quite late in human history. But here's a thought experiment. Return to our imaginary time machine (see above) and travel back to those civilizations that predate the invention of the numeral 0. How difficult would it be to teach those adult humans the position occupied by the symbol for zero?

[24] Daniel J. Povinelli and Sarah Dunphy-Lelii, "Do Chimpanzees Seek Explanations? Preliminary Comparative Investigations," *Canadian Journal of Comparative Psychology* 55 (2001): 187 – 195.

Patricia Smith Churchland

How do neurons know?

My knowing *anything* depends on my neurons – the cells of my brain.[1] More precisely, what I know depends on the specific configuration of connections among my trillion neurons, on the neurochemical interactions between connected neurons, and on the response portfolio of different neuron types. All this is what makes me *me*.

The range of things I know is as diverse as the range of stuff at a yard sale. Some is knowledge how, some knowledge that, some a bit of both, and some not exactly either. Some is fleeting, some enduring. Some I can articulate, such as the instructions for changing a tire, some, such as how I construct a logical argument, I cannot.

Some learning is conscious, some not. To learn some things, such as how to ride a bicycle, I have to try over and over; by contrast, learning to avoid eating oysters if they made me vomit the last time just happens. Knowing how to change a tire depends on cultural artifacts, but knowing how to clap does not.

And *neurons* are at the bottom of it all. How did it come to pass that we know *anything*?

Early in the history of living things, evolution stumbled upon the advantages accruing to animals whose nervous systems could make predictions based upon past correlations. Unlike plants, who have to take what comes, animals are movers, and having a brain that can learn confers a competitive advantage in finding food, mates, and shelter and in avoiding dangers. Nervous systems earn their keep in the service of prediction, and, to that end, map the *me-relevant* parts of the world – its spatial relations, social relations, dangers, and so on. And, of course, brains map their worlds in varying degrees of complexity, and relative to the needs, equipment, and lifestyle of the organisms they inhabit.[2]

Patricia Smith Churchland is UC *President's Professor of Philosophy and chair of the philosophy department at the University of California, San Diego, and adjunct professor at the Salk Institute. She is past president of the American Philosophical Association and the Society for Philosophy and Psychology. Her latest books are "Brain-Wise: Studies in Neurophilosophy" (2002) and "On the Contrary: Critical Essays, 1987 – 1997" (with Paul Churchland, 1998).*

[1] Portions of this paper are drawn from my book *Brain-Wise: Studies in Neurophilosophy* (Cambridge, Mass.: MIT Press, 2002).

[2] See Patricia Smith Churchland and Paul M. Churchland, "Neural Worlds and Real Worlds," *Nature Reviews Neuroscience* 3 (11) (November 2002): 903 – 907.

Thus humans, dogs, and frogs will represent the same pond quite differently. The human, for example, may be interested in the pond's water source, the potability of the water, or the potential for irrigation. The dog may be interested in a cool swim and a good drink, and the frog, in a good place to lay eggs, find flies, bask in the sun, or hide.

Boiled down to essentials, the main problems for the neuroscience of knowledge are these: How do structural arrangements in neural tissue embody knowledge (the problem of representations)? How, as a result of the animal's experience, do neurons undergo changes in their structural features such that these changes constitute knowing something new (the problem of learning)? How is the genome organized so that the nervous system it builds is able to learn what it needs to learn?

The spectacular progress, during the last three or four decades, in genetics, psychology, neuroethology, neuroembryology, and neurobiology has given the problems of how brains represent and learn and get built an entirely new look. In the process, many revered paradigms have taken a pounding. From the ashes of the old verities is arising a very different framework for thinking about ourselves and how our brains make sense of the world.

Historically, philosophers have debated how much of what we know is based on instinct, and how much on experience. At one extreme, the rationalists argued that essentially all knowledge was innate. At the other, radical empiricists, impressed by infant modifiability and by the impact of culture, argued that all knowledge was acquired.

Knowledge displayed at birth is obviously likely to be innate. A normal neonate rat scrambles to the warmest place, latches its mouth onto a nipple, and begins to suck. A kitten thrown into the air rights itself and lands on its feet. A human neonate will imitate a facial expression, such as an outstuck tongue. But other knowledge, such as how to weave or make fire, is obviously learned postnatally.

Such contrasts have seemed to imply that everything we know is either caused by genes or caused by experience, where these categories are construed as exclusive and exhaustive. But recent discoveries in molecular biology, neuroembryology, and neurobiology have demolished this sharp distinction between nature and nurture. One such discovery is that normal development, right from the earliest stages, relies on both genes and epigenetic conditions. For example, a female (XX) fetus developing in a uterine environment that is unusually high in androgens may be born with male-looking genitalia and may have a masculinized area in the hypothalamus, a sexually dimorphic brain region. In mice, the gender of adjacent siblings on the placental fetus line in the uterus will affect such things as the male/female ratio of a given mouse's subsequent offspring, and even the longevity of those offspring.

On the other hand, paradigmatic instances of long-term learning, such as memorizing a route through a forest, rely on genes to produce changes in cells that embody that learning. If you experience a new kind of sensorimotor event during the day – say, for example, you learn to cast a fishing line – and your brain rehearses that event during your deep sleep cycle, then the gene *zif*-268 will be up-regulated. Improvement in casting the next day will depend on the resulting gene products and their role in neuronal function.

Indeed, five important and related discoveries have made it increasingly clear just how interrelated 'nature' and 'nurture' are, and, consequently, how inadequate the old distinction is.[3]

First, what genes do is code for proteins. Strictly speaking, there is no gene for a sucking reflex, let alone for female coyness or Scottish thriftiness or cognizance of the concept of zero. A gene is simply a sequence of base pairs containing the information that allows RNA to string together a sequence of amino acids to constitute a protein. (This gene is said to be 'expressed' when it is transcribed into RNA products, some of which, in turn, are translated into proteins.)

Second, natural selection cannot directly select particular wiring to support a particular domain of knowledge. Blind luck aside, what determines whether the animal survives is its behavior; its equipment, neural and otherwise, underpins that behavior. Representational prowess in a nervous system can be selected for, albeit indirectly, only if the representational package informing the behavior was what gave the animal the competitive edge. Hence representational sophistication and its wiring infrastructure can be selected for only via the behavior they upgrade.

Third, there is a truly stunning degree of conservation in structures and developmental organization across all vertebrate animals, and a very high degree of conservation in basic cellular functions across phyla, from worms to spiders to humans. All nervous systems use essentially the same neurochemicals, and their neurons work in essentially the same way, the variations being vastly outweighed by the similarities. Humans have only about thirty thousand genes, and we differ from mice in only about three hundred of those;[4] meanwhile, we share about 99.7 percent of our genes with chimpanzees. Our brains and those of other primates have the same organization, the same gross structures in roughly the same proportions, the same neuron types, and, so far as we know, much the same developmental schedule and patterns of connectivity.

Fourth, given the high degree of conservation, whence the diversity of multicellular organisms? Molecular biologists have discovered that some genes regulate the expression of other genes, and are themselves regulated by yet other genes, in an intricate, interactive, and systematic organization. But genes (via RNA) make proteins, so the expression of one gene by another may be affected via sensitivity to protein products. Additionally, proteins, both within cells and in the extracellular space, may interact with each other to yield further contingencies that can figure in an unfolding regulatory cascade. Small differences in regulatory genes can have large and far-reaching effects, owing to the intricate hierarchy of regulatory linkages between them. The emergence of complex, interactive cause-effect profiles for gene expression begets very fancy regulatory cascades that can beget very fancy organisms – us, for example.

Fifth, various aspects of the development of an organism from fertilized egg to up-and-running critter depend on where and when cells are born. Neurons originate from the daughter cells of the last division of pre-neuron cells. Whether such a daughter cell becomes a glial (supporting) cell or a neuron, and which type of some hundred types of neurons

[3] In this discussion, I am greatly indebted to Barbara Finlay, Richard Darlington, and Nicholas Nicastro, "Developmental Structure in Brain Evolution," *Behavioral and Brain Sciences* 24 (2) (April 2001): 263 – 278.

[4] See John Gerhart and Marc Kirschner, *Cells, Embryos, and Evolution* (Oxford: Blackwell, 1997).

the cell becomes, depends on its epigenetic circumstances. Moreover, the manner in which neurons from one area, such as the thalamus, connect to cells in the cortex depends very much on epigenetic circumstances, e.g., on the spontaneous activity, and later, the experience-driven activity, of the thalamic and cortical neurons. This is not to say that there are no causally significant differences between, for instance, the neonatal sucking reflex and knowing how to make a fire. Differences, obviously, there are. The essential point is that the differences do not sort themselves into the archaic 'nature' versus 'nurture' bins. Genes and extragenetic factors collaborate in a complex interdependency.[5]

Recent discoveries in neuropsychology point in this same direction. Hitherto, it was assumed that brain centers – modules dedicated to a specific task – were wired up at birth. The idea was that we were able to see because dedicated 'visual modules' in the cortex were wired for vision; we could feel because dedicated modules in the cortex were wired for touch, and so on.

The truth turns out to be much more puzzling.

For example, the visual cortex of a blind subject is recruited during the reading of braille, a distinctly nonvisual, tactile skill – whether the subject has acquired or congenital blindness. It turns out, moreover, that stimulating the subject's visual cortex with a magnet-induced current will temporarily impede his braille performance. Even more remarkably, activity in the visual cortex occurs even in normal seeing subjects who are blindfolded for a few days while learning to read braille.[6] So long as the blindfold remains firmly in place to prevent any light from falling on the retina, performance of braille reading steadily improves. The blindfold is essential, for normal visual stimuli that activate the visual cortex in the normal way impede acquisition of the tactile skill. For example, if after five days the blindfold is removed, even briefly while the subject watches a television program before going to sleep, his braille performance under blindfold the next day falls from its previous level. If the visual cortex can be recruited in the processing of nonvisual signals, what sense can we make of the notion of the dedicated vision module, and of the dedicated-modules hypothesis more generally?

What is clear is that the nature versus nurture dichotomy is more of a liability than an asset in framing the inquiry into the origin of plasticity in human brains. Its inadequacy is rather like the inadequacy of 'good versus evil' as a framework for understanding the complexity of political life in human societies. It is not that there is nothing to it. But it is like using a grub hoe to remove a splinter.

An appealing idea is that if you learn something, such as how to tie a trucker's knot, then that information will be stored in one particular location in the brain, along with related knowledge – say, between reef knots and half-hitches. That is, after all, a good method for storing tools and paper files – in a particular drawer at a particular location. But this is not the brain's way, as Karl Lashley first demonstrated in the 1920s.

[5] See also Steven Quartz and Terrence J. Sejnowski, *Liars, Lovers, and Heroes* (New York: William Morrow, 2002).

[6] See Alvaro Pascual-Leone et al., "Study and Modulation of Human Cortical Excitability with Transcranial Magnetic Stimulation," *Journal of Clinical Neurophysiology* 15 (1998): 333 – 343.

Lashley reasoned that if a rat learned something, such as a route through a certain maze, and if that information was stored in a single, punctate location, then you should be able to extract it by lesioning the rat's brain in the right place. Lashley trained twenty rats on his maze. Next he removed a different area of cortex from each animal, and allowed the rats time to recover. He then retested each one to see which lesion removed knowledge of the maze. Lashley discovered that a rat's knowledge could not be localized to any single region; it appeared that all of the rats were somewhat impaired and yet somewhat competent – although more extensive tissue removal produced more serious memory deficit.

As improved experimental protocols later showed, Lashley's non-localization conclusion was essentially correct. There is no such thing as a dedicated memory organ in the brain; information is not stored on the filing cabinet model at all, but distributed across neurons.

A general understanding of what it means for information to be distributed over neurons in a network has emerged from computer models. The basic idea is that artificial neurons in a network, by virtue of their connections to other artificial neurons and of the variable strengths of those connections, can produce a pattern that represents something – such as a male face or a female face, or the face of Churchill. The connection strengths vary as the artificial network goes through a training phase, during which it gets feedback about the adequacy of its representations given its input. But many details of how actual neural nets – as opposed to computer-simulated ones – store and distribute information have not yet been pinned down, and so computer models and neural experiments are coevolving.

Neuroscientists are trying to understand the structure of learning by using a variety of research strategies. One strategy consists of tracking down experience-dependent changes at the level of the neuron to find out what precisely changes, when, and why. Another strategy involves learning on a larger scale: what happens in behavior and in particular brain subsystems when there are lesions, or during development, or when the subject performs a memory task while in a scanner, or, in the case of experimental animals, when certain genes are knocked out? At this level of inquiry, psychology, neuroscience, and molecular biology closely interact.

Network-level research aims to straddle the gap between the systems and the neuronal levels. One challenge is to understand how distinct local changes in many different neurons yield a coherent global, system-level change and a task-suitable modification of behavior. How do diverse and far-flung changes in the brain underlie an improved golf swing or a better knowledge of quantum mechanics?

What kinds of experience-dependent modifications occur in the brain? From one day to the next, the neurons that collectively make me what I am undergo many structural changes: new branches can sprout, existing branches can extend, and new receptor sites for neurochemical signals can come into being. On the other hand, pruning could decrease branches, and therewith decrease the number of synaptic connections between neurons. Or the synapses on remaining branches could be shut down altogether. Or the whole cell might die, taking with it all the synapses it formerly supported. Or, finally, in certain special regions, a whole new neuron might be born and begin to establish synaptic connections in its region.

And that is not all. Repeated high rates of synaptic firing (spiking) will deplete the neurotransmitter vesicles available for release, thus constituting a kind of memory on the order of two to three seconds. The constituents of particular neurons, the number of vesicles released per spike, and the number of transmitter molecules contained in each vesicle, can change. And yet, somehow, my skills remain much the same, and my autobiographical memories remain intact, even though my brain is never exactly the same from day to day, or even from minute to minute.

No 'bandleader' neurons exist to ensure that diverse changes within neurons and across neuronal populations are properly orchestrated and collectively reflect the lessons of experience. Nevertheless, several general assumptions guide research. For convenience, the broad range of neuronal modifiability can be condensed by referring simply to the modification of synapses. The decision to modify synapses can be made either globally (broadcast widely) or locally (targeting specific synapses). If made globally, then the signal for change will be permissive, in effect saying, "You may change yourself now" – but not dictating exactly where or by how much or in what direction. If local, the decision will likely conform to a rule such as this: If distinct but simultaneous input signals cause the receiving neuron to respond with a spike, then strengthen the connection between the input neurons and the output neurons. On its own, a signal from one presynaptic (sending) neuron is unlikely to cause the postsynaptic (receiving) neuron to spike. But if two distinct presynaptic neurons – perhaps one from the auditory system and one from the somatosensory system – connect to the same postsynaptic neuron at the same time, then the receiving neuron is more likely to spike. This joint input activity creates a larger postsynaptic effect, triggering a cascade of events inside the neuron that strengthens the synapse. This general arrangement allows for distinct but associated world events (e.g., blue flower and plenty of nectar) to be modeled by associated neuronal events.

The nervous system enables animals to make predictions.[7] Unlike plants, animals can use past correlations between classes of events (e.g., between red cherries and a satisfying taste) to judge the probability of future correlations. A central part of learning thus involves computing which specific properties predict the presence of which desirable effects. We correlate variable rewards with a feature to some degree of probability, so good predictions will reflect both the expected value of the reward and the probability of the reward's occurring; this is the expected utility. Humans and bees alike, in the normal course of the business of life, compute expected utility, and some neuronal details are beginning to emerge to explain how our brains do this.

To the casual observer, bees seem to visit flowers for nectar on a willy-nilly basis. Closer observation, however, reveals that they forage methodically. Not only do bees tend to remember which individual flowers they have already visited, but in a field of mixed flowers with varying amounts of nectar they also learn to optimize their foraging strategy, so that they get the most nectar for the least effort.

Suppose you stock a small field with two sets of plastic flowers – yellow and blue – each with wells in the center into which precise amounts of sucrose have

[7] John Morgan Allman, *Evolving Brains* (New York: Scientific American Library, 1999).

been deposited.[8] These flowers are randomly distributed around the enclosed field and then baited with measured volumes of 'nectar': all blue flowers have two milliliters; one-third of the yellow flowers have six milliliters, two-thirds have none. This sucrose distribution ensures that the mean value of visiting a population of blue flowers is the same as that of visiting the yellow flowers, though the yellow flowers are more uncertain than the blues.

After an initial random sampling of the flowers, the bees quickly fall into a pattern of going to the blue flowers 85 percent of the time. You can change their foraging pattern by raising the mean value of the yellow flowers – for example, by baiting one-third of them with ten milliliters. The behavior of the bees displays a kind of trade-off between the reliability of the source type and the nectar volume of the source type, with the bees showing a mild preference for reliability. What is interesting is this: depending on the reward profile taken in a sample of visits, the bees revise their strategy. The bees appear to be calculating expected utility. How do bees – *mere* bees – do this?

In the bee brain there is a neuron, though itself neither sensory nor motor, that responds positively to reward. This neuron, called VUMmx1 ('vum' for short), projects very diffusely in the bee brain, reaching both sensory and motor regions, as it mediates reinforcement learning. Using an artificial neural network, Read Montague and Peter Dayan discovered that the activity of vum represents prediction error – that is, the difference between 'the goodies expected' and 'the goodies received this time.'[9] Vum's output is the release of a neuromodulator that targets a variety of cells, including those responsible for action selection. If that neuromodulator also acts on the synapses connecting the sensory neurons to vum, then the synapses will get stronger, depending on whether the vum calculates 'worse than expected' (less neuromodulator) or 'better than expected' (more neuromodulator). Assuming that the Montague-Dayan model is correct, then a surprisingly simple circuit, operating according to a fairly simple weight-modification algorithm, underlies the bee's adaptability to foraging conditions.

Dependency relations between phenomena can be very complex. In much of life, dependencies are conditional and probabilistic: *If* I put a fresh worm on the hook, and *if* it is early afternoon, then *very probably* I will catch a trout *here*. As we learn more about the complexities of the world, we 'upgrade' our representations of dependency relations;[10] we learn, for example, that trout are more likely to be caught when the water is cool, that shadowy pools are more promising fish havens than sunny pools, and that talking to the worm, entreating the trout, or wearing a 'lucky' hat makes no difference. Part of what we call intelligence in humans and other animals is the capacity to acquire an increasingly complex understanding of dependency relations. This allows us to distinguish

[8] This experiment was done by Leslie Real, "Animal Choice Behavior and the Evolution of Cognitive Architecture," *Science* (1991): 980–986.

[9] See Read Montague and Peter Dayan, "Neurobiological Modeling," in William Bechtel, George Graham, and D. A. Balota, eds., *A Companion to Cognitive Science* (Malden, Mass.: Blackwell, 1998).

[10] Clark N. Glymour, *The Mind's Arrows* (Cambridge, Mass.: MIT Press, 2001). See also Alison Gopnik, Andrew N. Meltzoff, and Patricia K. Kuhl, *The Scientist in the Crib* (New York: William Morrow & Co., 1999).

fortuitous correlations that are not genuinely predictive in the long run (e.g., breaking a tooth on Friday the thirteenth) from causal correlations that are (e.g., breaking a tooth and chewing hard candy). This means that we can replace superstitious hypotheses with those that pass empirical muster.

Like the bee, humans and other animals have a reward system that mediates learning about how the world works. There are neurons in the mammalian brain that, like vum, respond to reward.[11] They shift their responsiveness to a stimulus that predicts reward, or indicates error if the reward is not forthcoming. These neurons project from a brainstem structure (the ventral tegmental area, or 'VTA') to the frontal cortex, and release dopamine onto the postsynaptic neurons. The dopamine, only one of the neurochemicals involved in the reward system, modulates the excitability of the target neurons to the neurotransmitters, thus setting up the conditions for local learning of specific associations.

Reinforcing a behavior by increasing pleasure and decreasing anxiety and pain works very efficiently. Nevertheless, such a system can be hijacked by plant-derived molecules whose behavior mimics the brain's own reward system neurochemicals. Changes in reward system pathways occur after administration of cocaine, nicotine, or opiates, all of which bind to receptor sites on neurons and are similar to the brain's own peptides. The precise role in brain function of the large number of brain peptides is one of neuroscience's continuing conundrums.[12]

These discoveries open the door to understanding the neural organization underlying prediction. They begin to forge the explanatory bridge between experience-dependent changes in single neurons and experience-dependent guidance of behavior. And they have begun to expose the neurobiology of addiction. A complementary line of research, meanwhile, is untangling the mechanisms for predicting what is nasty. Although aversive learning depends upon a different set of structures and networks than does reinforcement learning, here too the critical modifications happen at the level of individual neurons, and these local modifications are coordinated across neuronal populations and integrated across time.

Within other areas of learning research, comparable explanatory threads are beginning to tie together the many levels of nervous system organization. This research has deepened our understanding of working memory (holding information at the ready during the absence of relevant stimuli) spatial learning, autobiographical memory, motor skills, and logical inference. Granting the extraordinary research accomplishments in the neuroscience of knowledge, nevertheless it is vital to realize that these are still very early days for neuroscience. Many surprises – and even a revolution or two – are undoubtedly in store.

Together, neuroscience, psychology, embryology, and molecular biology are teaching us about ourselves as *knowers* – about what it is to know, learn, remember, and forget. But not all philosophers embrace these developments as progress.[13] Some believe that what we call

[11] See Paul W. Glimcher, *Decisions, Uncertainty, and the Brain* (Cambridge, Mass.: MIT Press, 2003).

[12] I am grateful to Roger Guillemain for discussing this point with me.

[13] I take it as a sign of the backwardness of academic philosophy that one of its most esteemed living practitioners, Jerry Fodor, is widely sup-

external reality is naught but an idea created in a nonphysical mind, a mind that can be understood only through introspection and reflection. To these philosophers, developments in cognitive neuroscience seem, at best, irrelevant.

The element of truth in these philosophers' approach is their hunch that the mind is not just a passive canvas on which reality paints. Indeed, we know that brains are continually organizing, structuring, extracting, and creating. As a central part of their predictive functions, nervous systems are rigged to make a coherent story of whatever input they get. 'Coherencing,' as I call it, sometimes entails seeing a fragment as a whole, or a contour where none exists; sometimes it involves predicting the imminent perception of an object as yet unperceived. As a result of learning, brains come to recognize a stimulus as indicating the onset of meningitis in a child, or an eclipse of the Sun by the Earth's shadow. Such knowledge depends upon stacks upon stacks of neural networks. There is no apprehending the nature of reality except via brains, and via the theories and artifacts that brains devise and interpret.

From this it does not follow, however, that reality is *only* a mind-created idea. It means, rather, that our brains have to keep plugging along, trying to devise hypotheses that more accurately map the causal structure of reality. We build the next generation of theories upon the scaffolding – or the ruins – of the last. How do we know whether our hypotheses are increasingly adequate? Only by their relative success in predicting and explaining.

But does all of this mean that there is a kind of fatal circularity in neuroscience – that the brain necessarily uses itself to study itself? Not if you think about it. The brain I study is seldom my own, but that of other animals or humans, and I can reliably generalize to my own case. Neuroepistemology involves many brains – correcting each other, testing each other, and building models that can be rated as better or worse in characterizing the neural world.

Is there anything left for the philosopher to do? For the neurophilosopher, at least, questions abound: about the integration of distinct memory systems, the nature of representation, the nature of reasoning and rationality, how information is used to make decisions, what nervous systems interpret as information, and so on. These are questions with deep roots reaching back to the ancient Greeks, with ramifying branches extending throughout the history and philosophy of Western thought. They are questions where experiment and theoretical insight must jointly conspire, where creativity in experimental design and creativity in theoretical speculation must egg each other on to unforeseen discoveries.[14]

ported for the following conviction: "If you want to know about the mind, study the mind – not the brain, and certainly not the genes" (*Times Literary Supplement,* 16 May 2003, 1 – 2). If philosophy is to have a future, it will have to do better than that.

[14] Many thanks to Ed McAmis and Paul Churchland for their ideas and revisions.