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 co-author (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 re-create 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 oncom-

ing 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.