In J. Brockman (Ed.)(2002) The Next Fifty Years: Science in the 21st century. New York: Vintage.

What Children Will Teach Scientists

Alison Gopnik

In 1997 space scientists working for NASA figured out how to tell whether there had once been water on Mars, a condition for life, by analyzing the light reflected from Mars rocks. Water would leave traces of carbonate on the rocks and this would influence the spectrum of the light reflected off the rocks. The scientists could work backward from data about the light, to the carbonates, to the presence of water. That same year, a few miles away in a Berkeley preschool, a four-year-old boy named Kevin, with equal excitement, figured out how a new machine worked. When some combinations of blocks but not others, were on the machine, the machine would play music. Kevin worked backwards from this data to infer which blocks would make the machine go and he used this knowledge to make the machine play music himself. In the next fifty years, we will come to understand how both Kevin and the rocket scientists at NASA could make these amazing discoveries. The answer to that question will change the way we think about science, childhood, brains, and maybe even genes.

Human beings know an enormous amount about the world around them. We know about rocks, waves, and toaster ovens; rabbits, palm trees, and potted petunias; parents, children, and orthodontists--an innumerable, inexhaustible array of objects, plants, animals, and people. By and large, our knowledge is strikingly accurate: We make remarkably good predictions about how toaster ovens and petunias and orthodontists work, and we use those predictions every day when we push the "Bake" button or add more Miracle-Gro or make an appointment. We didn't know all this when we were born; somehow or other we learned it.

We also learn about matters beyond our everyday experience--Mars rocks, viruses, neurons. And that knowledge, too, is strikingly accurate--accurate enough to let us conquer or at least alleviate such ancient scourges as smallpox and depression, not to mention baldness, impotence, and migraine headaches.

But how is it that we know so much? After all, the only information that reaches us directly from the world is a pattern of infinitesimal photons hitting our retinas and disturbances of air vibrating at our eardrums. How is it possible to get from that limited and apparently incoherent information to the truth? "Truth" may seem like a grand and metaphysical notion, but we all know a multitude of everyday truths: heat makes bread toast, water makes plants grow, broken appointments make orthodontists irritable. From a psychological point of view, our knowledge of these truths is as remarkable and puzzling as our knowledge of the truths of theoretical physics or astronomy. How is it that a series of interactions between one type of physical object, a bag of skin with a brain at the top, and other physical objects, like toaster ovens or petunias or orthodontists, can lead one object to learn about the other?

The last fifty years of developmental psychology have made this question even more puzzling. New techniques have allowed us to understand more about children's minds than ever before. Babies and young children turn out to both know more and learn more than we would ever have thought possible. By the time they are three or four they have already learned fundamental facts about how the world works. A theory of learning has to explain how very small children, who can't yet read or write or even talk well, can learn so much so quickly. Our ability to learn can't just be due to education or training or elaborate social institutions--rather, it seems to be a fundamental part of our human nature.

In the past fifty years, cognitive science has told us a great deal about what our knowledge of the world is like, how we use that knowledge, and how that knowledge is encoded in our brains. Developmental cognitive science has also told us a lot about how our knowledge changes as we grow older. But we have not yet understood where that knowledge comes from or how it can give us a true picture of the world outside us. Learning has been relegated to the concluding "Unsolved Mysteries" chapter of cognitive science textbooks, right next to consciousness and romantic love. I'm not convinced that we'll understand consciousness much better in fifty years, and I'm even more skeptical about romantic love. But I do think that we will make real progress toward a scientific account of learning.

We can find a model for this account in another, apparently quite different area of cognitive science; human vision. Here is the problem of vision: Take the patterns of light that enter the eyes and turn that information into accurate representations of objects moving in space. How do we solve it? People seem to make some implicit and very general assumptions about how the light coming into their eyes is related to objects in space. For example, we seem to unconsciously assume that the light coming into the retina is a two-dimensional projection of a three-dimensional world, and we use this assumption to solve the problem of vision. We never think that we are living in Flatland--though logically we could be. In fact, babies seem to be born making this assumption; for instance, very young babies will pull back from an object that seems to be looming toward them.

But the really interesting thing is not so much that we know this fact about vision, but that assuming that this fact is true lets us discover an incredibly varied array of new facts. I make a completely unconscious general assumption that my retinal image is a 2-D projection of 3-D objects. This helps me to infer that the particular image on my retina at this moment must come from two discs connected by a thin rod, lying on the surface of the floor at a particularly weird angle. Knowing this fact in turn helps me solve the never-ending, ever-changing, practical challenge of finding my reading glasses.

Sometimes, of course, the assumptions may lead us astray, especially if a demonic psychologist has been at work inventing visual illusions. But much more often these assumptions are correct, and they let us draw the right conclusions about what the outside world is like.

But how can brains make assumptions? The assumptions I've been talking about translate into constraints on the output that a brain (or any other computer) produces when it receives certain inputs. When my retina fires in a particular way, only some neurons, and not others, will fire further downstream. Neuroscientists can record the output of particular cells in the visual cortex while an animal looks at something and so construct a kind of circuit diagram. The neurological work shows how these constraints work in practice, and how these computations are actually carried out in the brain.

In vision science, there has been a remarkable convergence of different disciplines. Psychologists tell us what kinds of representations of objects we construct from what kinds of visual information; they tell us what we perceive when a particular pattern of light hits our eyes. That defines the problem. Mathematicians demonstrate how it is possible to solve that problem by making certain very general assumptions about how objects and light are related. Computer scientists show how those solutions can be implemented as constraints on the operation of actual physical machines. And neuroscientists show how those solutions are implemented in the particular machines inside our skulls.

A similar strategy may help us understand how we learn: that is, define the problems that human children and adults solve, mathematically work out possible solutions to those problems given certain assumptions, see how those solutions can be implemented in machines, and, eventually, see how they are implemented in our brains. Recently there has been a similar convergence of new ideas about learning from different disciplines--the philosophy of science, artificial intelligence, statistics, and developmental psychology. In the next fifty years, this convergence could lead to a full-fledged scientific theory of how we learn.

Start out with the problem, or at least one problem. How do we learn about the causal structure of the world--about how things work and how one event makes other events happen? This is obviously an important problem in the practice of any kind of science, but it is also an important problem for even very young children. Developmental psychologists have demonstrated that children understand a lot about causal relationships. By three or four, children understand some of the same basic causal facts about toaster ovens and petunias and people that adults understand. They also know more at five than they do at three and less than they do at seven; like scientists, children seem to be good at learning new causal facts.

But causal knowledge is also one of the most notorious examples of the gap between what we experience and what we learn. The philosopher David Hume originally articulated the problem. All we see are contingencies between events. One type of event may always follow another type of event, but how do we know that one event caused the other? And in real life, causal relations rarely involve just two events; dozens of different events may be causally interrelated in complicated ways. In real life, it's actually unusual for one event always to follow another; moreover, we may not always know which of two events came first. This uncertainty and complexity makes even everyday causal problems complicated. Did the toaster-oven element smoke and burn the toast, or did the crumbs from the burning toast make the element smoke? Or did we set the temperature too high and independently burn the toast and cause the element to smoke? All we can see is the simultaneous mess.

Is there a way to sort out the mess? Intuitively, there are two things we can do. We can perform a series of experiments: For example, we could set the temperature knob up high without putting any bread in the oven, or we could scatter crumbs of burnt toast on the element while keeping the temperature low. If experiments are impossible, we can make careful observations to determine when the element smokes and when it doesn't. Does it smoke only when the temperature is high, whether or not there is burning toast in the oven? Or does it smoke only when there is burning toast, whether or not the temperature is high?

When we perform these experiments or make these observations, we are making assumptions about how the pattern of contingencies among a set of events is related to the causal relations among them, just as we make assumptions about how 2-D retinal images are related to 3-D objects. Just as we never think we live in Flatland, we never think we live in a world without causes. Of course, as with vision, a Humean demon could arrange the contingencies in a way that would fool us. But we progress by assuming that there are no such demons--that, to paraphrase Einstein, God is slick but not mean.

A group of philosophers of science at Carnegie-Mellon University, led by Clark Glymour, and the computer scientist Judea Pearl and his colleagues at UCLA, have started to develop a mathematical formalism that allows us to go beyond intuition and state these assumptions in a rigorous way. We can think about causal relations in terms of a formalism called a directed acyclic graph, often also called a Bayes net. These graphs tell us about how the presence of one variable (like the state of the toast) will influence another variable (like the state of the heating element). The basic assumption behind the formalism is that if one event causes another, then when the value of one variable changes, the value of the other variable will also be likely to change. If the crumbs cause the element to smoke, then the presence of the crumbs should make the presence of the smoke more likely. We can represent these causal relations as arrows connecting the variables. The Bayes-net formalism makes some simple and general assumptions abut how the pattern of causal relations--the pattern of arrows--is related to the patterns of contingency among the variables. Here are three different graphical representations of the relation between the temperature knob, the burning toast, and the smoking element, corresponding to the three causal hypotheses we have described:

A. Temperature knob > burning toast > smoking element

The temperature knob makes the toast burn, which makes the element smoke.

B. Temperature knob > smoking element > burning toast

The temperature knob makes the element smoke, which makes the toast burn.

C. Smoking element < temperature knob > burning toast

The temperature knob independently makes the element smoke and the toast burn.

Each of these causal structures has different implications for the patterns of contingency among the variables, given the basic assumptions about contingency and causality. This is what allows us to draw the right conclusions from our experiments and observations. For instance, if A is true and we remove the burning toast, we should no longer see any relation between the temperature knob and the element. If B or C is true, then we should still see such a relation. If we turn the temperature knob down and burn the toast independently, then the element will smoke if A is true but not if B or C is true. If we put the temperature on high but prevent the element from smoking, then the toast will burn if A or C is true but not if B is true. Similarly, different causal structures will lead us to observe different patterns of contingency among the variables, even if we don't do the experiments ourselves. The mathematical work lets us spell out all these connections between contingency and causation in detail, even with structures much more complicated than those I've described here.

This mathematical work provides us with a kind of causal logic. Classical deductive logic took off from a few basic assumptions about reasoning and, mathematically, turned those assumptions into a method for deriving true conclusions from true premises. The new causal work makes a few basic assumptions about causality and then provides a systematic method for deriving true conclusions about causal relationships from observation and experiment.

Computer scientists have started to turn this abstract mathematics into computer programs that can actually learn about the world. One of the biggest differences between computers and people has been that computer programs could do only what you told them to do in the first place. A real Turing test—a test that would tell if computers are like people–would require not only that a computer could do the same things as a human adult but also that it could learn how to do them based on the experiences of a human child.

Computer scientists translate the mathematical assumptions into constraints on the kinds of causal graphs a computer will produce when it is given certain patterns of contingency data. Using the new mathematical ideas, for example, computer scientists working for NASA designed programs that enable a robot to learn about the composition of rocks on Mars just by looking at data from a spectrometer, without having to consult experts back on Earth.

So far, all this may seem rather removed from the question we started out with: We wanted to know how ordinary people--and, in particular, ordinary children--actually learn, not just how high-powered scientists and statisticians and computers can learn. But there is just beginning to be evidence that all learners may make the same mathematical assumptions about the relation between causality and contingency. Psychologists who explore the ways in which ordinary adults figure out causal problems have independently hit on some of the same mathematical models as the investigators in philosophy.

Psychologists are beginning to find that children as young as two use this sort of causal logic as well. We can present children with the equivalent of the toaster oven--a machine that makes things happen in a somewhat complicated and mysterious way. Sometimes we give the children particular patterns of evidence about the machine, and sometimes we let the children do experiments to find that evidence themselves. Then we see if they can figure out how the machine works. Children are surprisingly good at drawing just the right causal conclusions from the data in precisely the ways that the formalism would predict--young children really are rocket scientists. Of course, unlike scientists, the children seem to be completely unconscious of how they are reaching their conclusions.

In the next fifty years, once we know just what the computations are that children and adults perform, we will be able to look into their brains and see how they perform them. As our brain-imaging techniques become more and more precise and we learn more and more about the computations, we will begin to see how our brains are designed to implement those computations. The answer to that question is likely to be related to the greatest breakthroughs to come in neuroscience. The most important thing about the brain is its ability to change in response to input from the environment, yet this is one of the aspects of the brain we know least about. It is as if we knew everything about the anatomical structure of dead hearts but almost nothing about how living hearts pump blood. Brains are, above all else, organs that learn, and if we know how learning works we will know something important about how brains work.

What's more, as the fifty years proceed, the answer to the problem of how minds and brains learn may also turn out to be related to an even more general developmental problem, that of morphological development. Another great unsolved problem of the next millennium is the question of how DNA instructions turn something as simple as a fertilized egg into something as gloriously complex as a newborn baby. One of the lessons of the past few years of genetic research is that the genome can't be a detailed set of instructions for creating an organism; it isn't a blueprint. But then how does it work? Genes seem to operate not so much by directly determining what a cell will do but by initiating a kind of causal cascade in the environment of a cell that ends up influencing the cell in predictable ways. For example, the genes that determine sexual morphology do so by producing testosterone, which then acts on the organism in a complex way. Occasionally the demons are unleashed, the environment turns out to be different than the one the genes "expected", and the system goes awry. But usually the environment is predictable, and the genome exploits that predictability to produce a complex organism.

It may be helpful to think of DNA instructions as encoding implicit general assumptions about the interaction between cells and their environments (largely other cells), both before birth and after it. In the psychological case, implicit assumptions about our relations to our environment let us build remarkably complex structures that are remarkably well adapted to that environment. It is at least conceivable that this same general approach may apply to the biological case.

The greatest achievement of a unified theory of learning, though, may be to demonstrate that the most brilliant scientists and the most ordinary kids are actually engaged in the same enterprise. At the end of the last century, knowledge began to become the most valuable currency, like land in a feudal economy or capital in an industrial economy. The new science of learning should tell us that knowledge is not just a prize to be won in some desperate test-taking struggle for places in the contemporary mandarinate. Instead it is, literally and not just rhetorically, our universal human birthright.

ALISON GOPNIK is a professor of psychology at the University of California at Berkeley. She is an international leader in the field of children's learning and was one of the first cognitive scientists to show how developmental psychology could help solve ancient philosophical problems. She is the coauthor (with Andrew Meltzoff) of Words, Thoughts, and Theories, and (with Patricia Kuhl and Andrew Meltzoff) of The Scientist in the Crib: Minds, Brains, and How Children Learn.