

JUMPING TO CONCLUSIONS

Can people be counted on to make sound judgments?

By
**Deanna
Kuhn**

A four-year-old watches as a monkey hand puppet approaches a vase containing a red and a blue plastic flower. The monkey sneezes. The monkey backs away, returns to sniff again, and again sneezes. An adult then removes the red flower and replaces it with a yellow one. The monkey comes up to smell the yellow and blue flowers twice and each time sneezes. The adult next replaces the blue flower with the red one. The monkey comes up to smell the red and yellow flowers and this time does not sneeze.

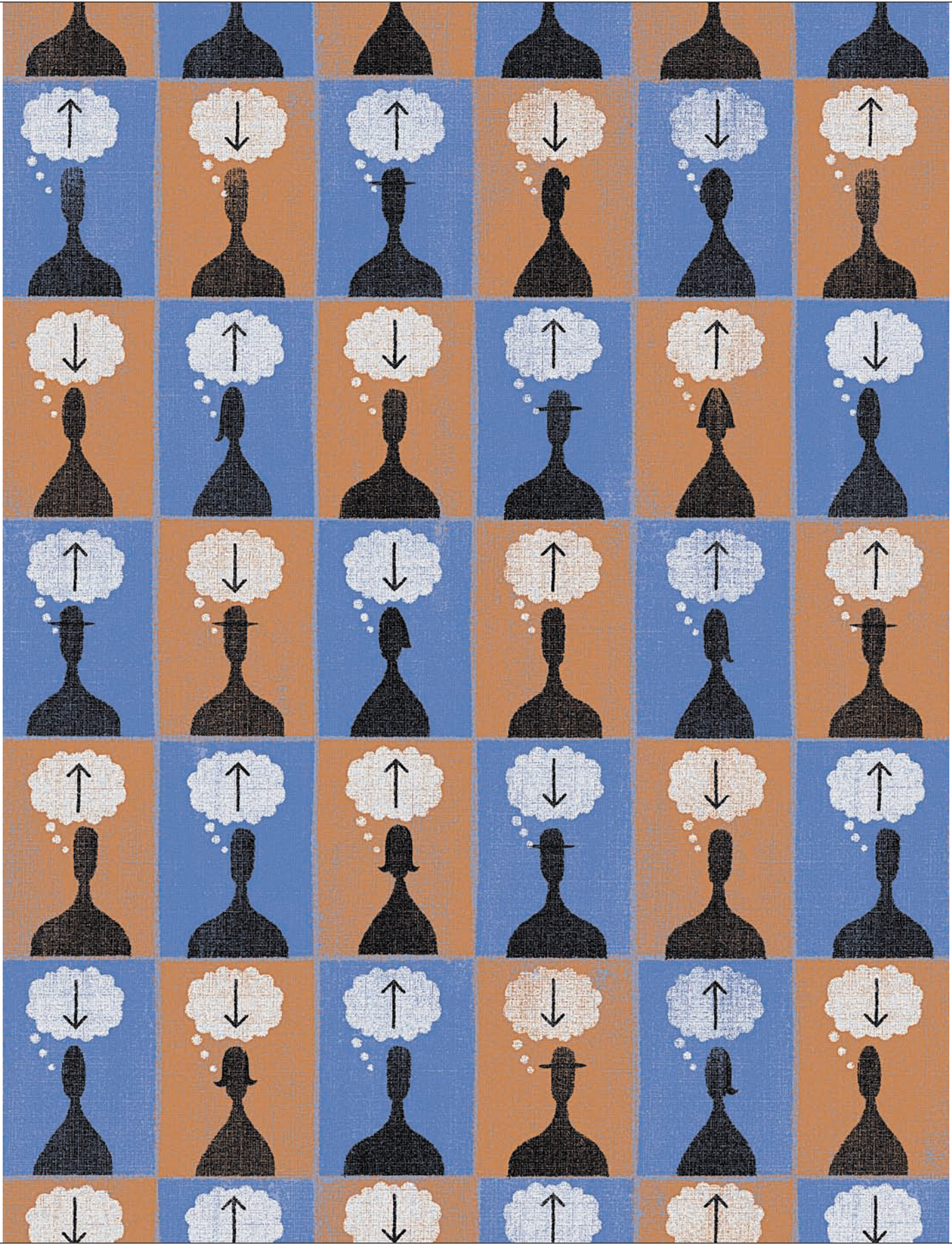
The child is then asked, “Can you give me the flower that makes Monkey sneeze?” When psychologists Laura E. Schulz and Alison Gopnik, both then at the University of California, Berkeley, did this experiment, 79 percent of four-year-olds correctly chose the blue flower. As their research makes clear, even very young children have begun to understand cause and effect. This process is critical to their ability to make sense of their world and to make their way in it.

With such powers of discernment already

in place by age four, people should be highly skilled at identifying cause and effect—causal reasoning—by the time they are adults, shouldn’t they? Indeed, a substantial body of contemporary research suggests that is the case, highlighting the nuanced judgments adults are capable of—such as making consistent estimates, across different circumstances, of the numerical probabilities that two events are causally related.

Here I present some evidence that gives a very different impression: the everyday causal reasoning of the average adult regarding familiar topics appears highly fallible. People connect two events as cause and effect based on little or no evidence, and they act on these judgments—they jump to conclusions. By learning more about precisely how they do so, researchers can develop ways to improve thinking. Such efforts could help educators in their mission to inspire solid, careful thinking in young minds.

A possible explanation for the discrepancy between our findings and much of the relevant literature is that researchers studying



Case Studies in Cause and Effect



In a study, adults had difficulty judging which factors yielded the best performance at fund-raising parties. The results provided (*below*) showed a causal relation between “auction” and sales (compare first and third parties) and no causal relation between “comedian” and sales (compare second

and third). Information was insufficient to determine other feature effects. Yet 83 percent of the volunteers said two or more features increased sales, and 45 percent claimed three or all four did so. Most also reported feeling certain about the correctness of their (often erroneous) judgments.

First party

- Door prizes
- Comedian
- Costumes

SALES: MEDIUM

Second party

- Door prizes
- Auction
- Costumes

SALES: HIGH

Third party

- Door prizes
- Auction
- Comedian
- Costumes

SALES: HIGH

A second reasoning task asked volunteers to make predictions, all of which were indeterminate (because the effects of door prizes and costumes were unknown). Respondents displayed inconsistent logic. Particularly difficult for them was recognizing that a feature whose presence had a positive influence on an outcome would negatively affect the

outcome when it was removed. For example, in the prediction question involving door prizes and a comedian (*lower left*), only 40 percent of respondents circled the absence of an auction as affecting the outcome, although 85 percent had correctly labeled it as causal. As before, people were nonetheless certain about their judgments. —D.K.

• Door prizes • Auction
• Comedian

SALES: LOW MEDIUM HIGH

How certain are you? (circle one)

Very certain Certain
Think so but not certain Just guessing

Which influenced your prediction? (circle as many as apply)

- Door prizes
- Auction
- Comedian
- Absence of costumes

• Auction • Costumes
• Comedian

SALES: LOW MEDIUM HIGH

How certain are you? (circle one)

Very certain Certain
Think so but not certain Just guessing

Which influenced your prediction? (circle as many as apply)

- Auction
- Costumes
- Comedian
- Absence of door prizes

• Door prizes
• Comedian

SALES: LOW MEDIUM HIGH

How certain are you? (circle one)

Very certain Certain
Think so but not certain Just guessing

Which influenced your prediction? (circle as many as apply)

- Door prizes
- Comedian
- Absence of auction
- Absence of costumes

• Auction
• Costumes

SALES: LOW MEDIUM HIGH

How certain are you? (circle one)

Very certain Certain
Think so but not certain Just guessing

Which influenced your prediction? (circle as many as apply)

- Auction
- Costumes
- Absence of comedian
- Absence of door prizes

JIM FRAZIER Getty Images

causal reasoning skills in adults have typically based their conclusions on studies of a narrow segment of the adult population in a specific context—college students in laboratory settings performing complex paper-and-pencil tasks. In a 2000 study, for example, psychologists Patricia Cheng of the University of California, Los Ange-

paring the second and third parties, you can see that adding a comedian has no effect on sales. Yet the information available is insufficient for assessing the causal status of door prizes or costumes (because they are always present).

Did this diverse group of adults at Penn Station show as much skill in isolating cause and ef-

Do studies represent the cognitive performance of average people in their thinking about everyday affairs?

les, and Yunnwen Lien of the National Taiwan University in Taipei presented college students with a set of instances that described the blooming frequencies of plants that had been fed plant food of different shapes and colors. After examining each case, students rated on a numerical scale the likelihood or degree of causal influence of each of the factors and/or made predictions regarding outcomes for novel instances—and showed good reliability in doing so. Although such studies highlight the skills that college students display in such tasks, do they represent the cognitive performance of average people in their thinking about everyday affairs?

To address this question, my student Joanna Saab and I went last year to New York City’s Pennsylvania Station. We asked 40 people seated in the waiting room if they would spend 10 minutes answering a survey in exchange for five dollars. Virtually all accepted. We explained that a group was trying different combinations of entertainment features at fund-raisers, to see which would sell the most tickets, and showed each person a diagram with some of the results. The sign for the first party listed door prizes, comedian, costumes; its sales were “medium.” The second party listed door prizes, auction, costumes; its sales were “high.” The third party listed door prizes, auction, comedian, costumes; its sales were “high.”

We left the diagram in view as we talked to each of our interviewees, and we asked, “Based on their results, does the auction help ticket sales?” We also asked how certain they were about their answers. They could choose “very certain,” “certain,” “think so but not certain” or “just guessing.” We asked the same questions for each of the three remaining features: comedian, door prizes, costumes.

As you can deduce for yourself [see box on opposite page], if you examine the first and third parties, adding the auction boosts sales. By com-

paring the second and third parties, you can see that adding a comedian has no effect on sales. Yet the information available is insufficient for assessing the causal status of door prizes or costumes (because they are always present). Even more striking, most respondents were quite confident that they were correct. For two of the four features, the average certitude reported was greater than “certain” (and tending toward “very certain”), whereas for the other two the average was slightly below “certain.” Gender was not a factor: men and women did not differ significantly in either their judgments or levels of certainty.

What made these respondents so sure about which features affected outcome and which did not? We emphasized to them that they should base their conclusions on the results shown for the particular group of people indicated (rather than on their own prior beliefs about the effectiveness of these features); in response to a follow-up query at the end, all respondents indicated that they had done so. Yet their responses revealed that their judgments were in fact influenced by their own ideas about how effective these features ought to be. Respondents judged door prizes to affect outcome (83 percent did so) much more commonly than they judged costumes to affect outcome (33 percent did so), although the evidence with respect to the two features was identical.

To gain further insight, we presented respondents with an additional task [see box on opposite page].

In this second case, there were no correct answers. One cannot make justifiable predictions

given the indeterminate causal status of two of the features: door prizes and costumes. Nevertheless, respondents' certainty regarding the predictions they made remained as high as it had been for their causal judgments. Their predictions, moreover, were informative. For example, to infer whether a respondent judged the auction feature as causal, we compared the predictions the person made for a particular pair of cases—specifically, those two cases that involved door prizes. If the auction was being regarded as causal, predictions for these two cases (one with the auction present and the other with it absent) should have differed. If the auction was being regarded as noncausal, its presence or absence should have had no influ-

motivated to answer the questions to the best of their ability to justify receiving their five dollars. But they were unlikely to focus on the task as a reasoning test, designed to assess their mental processes, as readily as would college students, who have become familiar with such tests. The purpose, which most college students recognize, is not to achieve a solution (whether it be maximizing ticket sales or designing a bridge sufficient to support a given weight) but rather to display how they go about tackling the problem. College students have learned to behave accordingly, looking at the information given and determining how they should use it to produce an answer. Unsurprisingly, then, we found that re-

Aren't people most likely to come to the best conclusions if they make use of all they know?

ence and predictions for these two cases should have been identical. Similarly, comparing the predictions for the two cases involving costumes allowed us to infer whether the respondent judged the comedian as causal.

The implicit judgments that respondents made in the prediction task tended to be inconsistent with the causal judgments they had made in the judgment task when they were asked to indicate explicitly whether a factor was causal (“helped ticket sales”). Only 15 percent made consistent judgments across both tasks. Similarly, people were inconsistent in the implicit causal attributions they made in response to the questions about which features had influenced each of their predictions. Among the 63 percent who had correctly judged the inclusion of a comedian as having no causal effect in the judgment task, for example, a majority nonetheless indicated that the presence or absence of a comedian had influenced their predictions. Particularly difficult was recognizing that a feature whose presence positively affected an outcome would negatively affect the outcome when it was removed.

Reconciling the Inconsistencies

How can we reconcile the inconsistent and incautious causal judgments made by people waiting in a train station—judgments they claimed to be certain of—with the reasoning skills observed in college students and even four-year-olds? The answer is invariably multifaceted. Our respondents took the task seriously and were

spondents with a college background made sounder judgments than those without it did.

Those who do not possess this “academic” mind-set, in contrast, tend to focus on getting the problem solved and allocate little attention to the mental operations they use in the process. In getting to a solution, they bring to bear everything they know that might be of use. Based on their own prior knowledge that door prizes seem more likely to be a winner for fund-raising than costumes, they judge door prizes as causal—even though the presented evidence provides no support for this difference. Keeping track of how they responded in an earlier part of the interview, so as to maintain consistency, will not help solve the problem and thus is not a high priority. For such people, the best reading of how things look at the moment is what is important. Once a decision is reached, moreover, expressing confidence and certainty is better than wavering.

So who is using the “smarter” approach? Why put old beliefs on hold when evaluating new information? Aren't people most likely to come to the best conclusions if they make use of all they know while reaching them? In many contexts, the answer is yes. Yet being able to evaluate “the information given” to determine exactly what it does (and does not) imply is also an important skill—and not just within the rarefied halls of academia.

Suppose, for example, I am thinking about trying the new weight-loss product my friends are talking about, but they tell me they have



heard it could cause cancer. When I go to the medical library to look up a recent study on the product, I want to be able to interpret what it says, independent of prior thoughts I may have. In reaching a decision, I may ultimately integrate what the report says with other considerations. But I could not do so were I not able to interpret the document in its own right.

In his 2004 book, *The Robot's Rebellion*, Keith E. Stanovich of the University of Toronto similarly makes the case for the importance of what he calls “decontextualized” reasoning and describes studies in which participants fail to use it. The relevance of such reasoning is by no means limited to thinking about causality. Reaching a verdict in a legal trial, for example, is one common context in which jurors are required to rely on the presented evidence alone, not on everything that comes to mind related to this evidence. So is deductive reasoning, employing ancient Greek philosopher Aristotle’s classical syllogisms. Stanovich notes, for example, that 70

percent of adult subjects accepted this syllogism as valid:

Premise 1: All living things need water.

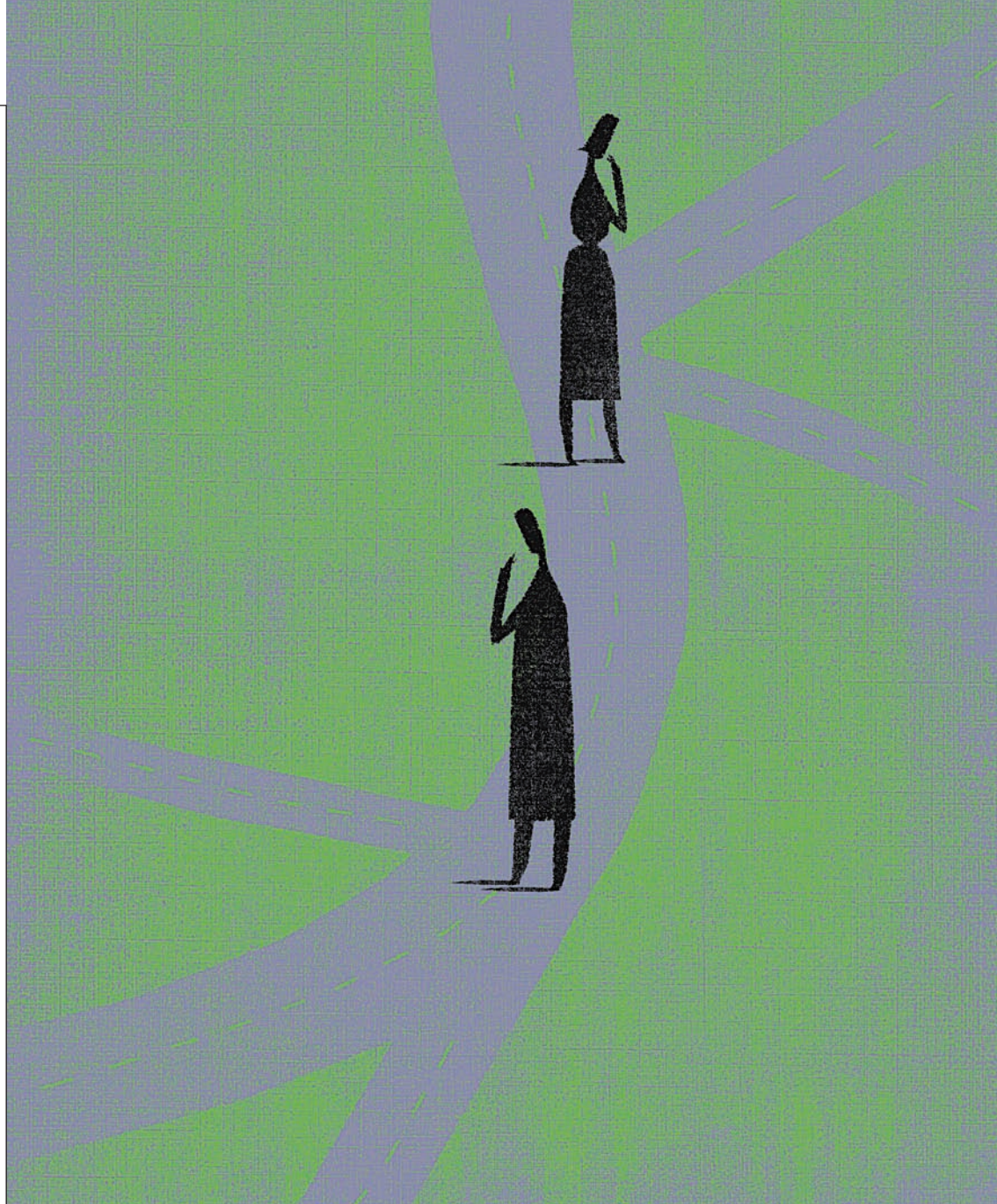
Premise 2: Roses need water.

Conclusion: Roses are living things.

Because we know the conclusion to be true in the real world, it is easy to accept, even though it

(The Author)

DEANNA KUHN is professor of psychology and education at Teachers College, Columbia University. She holds a Ph.D. in developmental psychology from the University of California, Berkeley, and was previously at the Graduate School of Education at Harvard University. She is editor of the journal *Cognitive Development* and co-editor of the *Cognition, Perception, and Language* volume of the *Handbook of Child Psychology*, sixth edition (Wiley, 2006). Kuhn is co-author of *The Development of Scientific Thinking Skills* (Academic Press, 1988). Her most recent book, *Education for Thinking* (Harvard University Press, 2005), seeks to identify and examine the thinking skills that we can justify as objectives of education.



does not follow logically from the premises. To be convinced of this fact, we need only compare it with a syllogism identical in form:

Premise 1: All animals of the hudon class are ferocious.

Premise 2: Wampets are ferocious.

Conclusion: Wampets are animals of the hudon class.

Typically only 20 percent of people accept this conclusion as correct. The other 80 percent correctly reject it, the improvement in performance presumably arising because no obfuscating real-world knowledge got in the way.

As the research we conducted at the train sta-

tion suggests, decontextualization is not the only skill in the careful reasoner's mental tool kit. Consistency and avoiding undue certainty in one's judgments are also important. Undue certainty reflects a failure in "knowing what you know" (also called metacognition) and underlies the rigidity in thinking that is a major contributor to human strife. Inconsistency can be similarly self-serving, allowing us to protect our favorite theories without subjecting them to the same standards of evidence to which we subject those of others. We maintain that superior skill was the cause of our team's victory, whereas the other team's win was because of luck.

The authors made no assessment of consistency or certainty of the causal judgments of the four-

year-olds in the study described earlier. But we can see why these children may have had an easier time evaluating evidence than the adults in our study had. The scenario involving different colored flowers engaged very little in the way of prior knowledge regarding which colors would be more likely to make a monkey sneeze. The adults, in contrast, had much prior experience that they could bring to bear on matters of event planning, ticket sales and the enjoyableness of different activities. This

nable to improvement, and with practice it becomes more careful and critical. Performance on standardized tests of “basic skills” of literacy and numeracy has come to occupy center stage as a measure of how successful schooling has been at teaching students what they need to know. In contrast, learning to make sound judgments about matters of the kind people encounter in everyday life has not been a high priority as an objective of education.

(Rich knowledge can make it more challenging to evaluate evidence in its own right.)

rich knowledge made it more challenging for them to evaluate the evidence in its own right.

What the competence displayed by the subjects in Schulz and Gopnik’s study does show, however, is that the underlying reasoning processes entailed in multivariable causal inference (involving multiple potential causes) have developed to at least a rudimentary degree among four-year-olds. More important, this is competence that we can build on in devising the kinds of educational experiences that will help older children and adolescents, and even adults, become more careful causal reasoners.

Other research that my colleagues and I have done shows that both children and adults do come to reason more critically about causality if they are provided frequent opportunities to practice evaluating evidence and making causal judgments and predictions. Early adolescent students initially show the kinds of faulty multivariable causal reasoning that have been illustrated here. But if they engage with problems of this kind over the course of several months, their reasoning improves sharply. The same is true of young adults enrolled in a community college.

Thinking Forward

The message we might glean from the research I have described is twofold. First, the causal reasoning of average adults regarding everyday matters is in fact highly fallible. People frequently make unwarranted inferences with unwarranted certainty, and it is likely that they act on many of these inferences.

Second, although people may leap to unwarranted conclusions in their judgments about causality, we should not jump to the conclusion that this is the way things must be. Thinking is ame-

Such aspects of cognition may be recognized as warranting more attention, as people today struggle to interpret escalating amounts of information about increasingly complex matters, some of which have implications for their very survival. By promoting the development of skills that will help them meet this challenge, we could enrich conceptions of what is important for students to learn. As noted earlier, frequent opportunity to investigate diverse forms of evidence and draw conclusions from them does strengthen reasoning skills. Even getting into the habit of asking oneself and others simple questions like “How do we know?” and “Can we be certain?” goes a long way toward the objective of sound, rigorous thinking.

In an era of escalating pressure on educators to produce the standardized test performance demanded by No Child Left Behind legislation, is it sensible for them to even think about undertaking anything more? Certainly young people must become literate and numerate. But in the end, what could be a more important purpose of education than to help students learn to exercise their minds to make the kinds of careful, thoughtful judgments that will serve them well over a lifetime? **M**

(Further Reading)

- ◆ **Who Is Rational? Studies of Individual Differences in Reasoning.** Keith E. Stanovich. Lawrence Erlbaum Associates, 1999.
- ◆ **The Everyday Production of Knowledge: Individual Differences in Epistemological Understanding and Juror Reasoning Skill.** M. Weinstock and M. A. Cronin in *Applied Cognitive Psychology*, Vol. 17, No. 2, pages 161–181; 2003.
- ◆ **The Robot’s Rebellion: Finding Meaning in the Age of Darwin.** Keith E. Stanovich. University of Chicago Press, 2004.
- ◆ **Education for Thinking.** Deanna Kuhn. Harvard University Press, 2005.