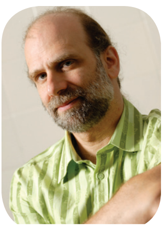
Detecting Cheaters



BRUCE SCHNEIER
BT

Our brains are specially designed to deal with cheating in social exchanges. The evolutionary psychology explanation is that we evolved brain heuristics for the social problems that our prehistoric ancestors had to deal with. Once humans became good at cheating, they then had to become good at detecting cheating—otherwise, the social group would fall apart.

Perhaps the most vivid demonstration of this can be seen with variations on what's known as the Wason selection task, named after the psychologist who first studied it. Back in the 1960s, it was a test of logical reasoning; today, it's used more as a demonstration of evolutionary psychology. But before we get to the experiment, let's get into the mathematical background.

Propositional calculus is a system for deducing conclusions from true premises. It uses variables for statements because the logic works regardless of what the statements are. College courses on the subject are taught by either the mathematics or the philosophy department, and they're not generally considered to be easy classes. Two particular rules of inference are relevant here: *modus ponens* and *modus tollens*. Both allow you to reason from a statement of the form, "if *P*, then *Q*." (If Socrates was a man, then Socrates was mortal. If you are to eat dessert, then you must first eat your vegetables. If it is raining, then Gwendolyn had Crunchy Wunchies for breakfast. That sort of thing.) *Modus ponens* goes like this:

> If *P*, then *Q*. *P*. Therefore, *Q*.

In other words, if you assume the conditional rule is true, and if you assume the antecedent of that rule is true, then the consequent is true. So,

> If Socrates was a man, then Socrates was mortal. Socrates was a man. Therefore, Socrates was mortal.

Modus tollens is more complicated:

> If *P*, then *Q*. Not *Q*. Therefore, not *P*.

> If Socrates was a man, then Socrates was mortal. Socrates was not mortal. Therefore, Socrates was not a man.

This makes sense: if Socrates was not mortal, then he was a demigod or a stone statue or something.

Both are valid forms of logical reasoning. If you know "if *P*, then *Q*" and "*P*," then you know "*Q*." If you know "if *P*, then *Q*" and "not *Q*," then you know "*P*." (The other two similar forms don't work. If you know "if *P*, then *Q*" and "*Q*," you don't know anything about "*P*." And if you know "if *P*, then *Q*" and "not *P*," then you don't know anything about "*Q*.")

If I explained this in front of an audience full of normal people, not mathematicians or philosophers, most of them would be lost. Unsurprisingly, they would have trouble either explaining the rules or using them properly. Just ask any grad student who has had to teach a formal logic class; people have trouble with this.

Consider the Wason selection task. Subjects are presented with four cards next to each other on a table. Each card represents a person, with each side listing some statement about that person. The subject is then given a general rule and asked which cards he would have to turn over to ensure that the four people satisfied that rule. For example, the general rule might be, "If a person travels to Boston, then he or she takes a plane." The four cards might correspond to travelers and have a destination on one side and a mode of transport on the other. On the side facing the subject, they read: "went to Boston," "went to New York," "took a plane," and "took a car." Formal logic states that the rule is violated if someone goes to Boston without taking a plane. Translating into propositional calculus, there's the general rule: if *P*, then *Q*. The four cards are "*P*," "not *P*," "*Q*," and "not *Q*." To verify that "if *P*,

cont. on p. 97

 1540-7993/11/$26.00 © 2011 IEEE

cont. from p. 96

then *Q*" is a valid rule, you have to verify *modus ponens* by turning over the "*P*" card and making sure that the reverse says "*Q*." To verify *modus tollens*, you turn over the "not *Q*" card and make sure that the reverse doesn't say "*P*."

Shifting back to the example, you need to turn over the "went to Boston" card to make sure that person took a plane, and you need to turn over the "took a car" card to make sure that person didn't go to Boston. You don't—as many people think—need to turn over the "took a plane" card to see if it says "went to Boston" because you don't care. The person might have been flying to Boston, New York, San Francisco, or London. The rule only says that people going to Boston fly; it doesn't break the rule if someone flies elsewhere.

If you're confused, you aren't alone. When Wason first did this study, fewer than 10 percent of his subjects got it right. Others replicated the study and got similar results. The best result I've seen is "fewer than 25 percent." Training in formal logic doesn't seem to help very much. Neither does ensuring that the example is drawn from events and topics with which the subjects are familiar. People are just bad at the Wason selection task. They also tend to only take college logic classes upon requirement.

This isn't just another "math is hard" story. There's a point to this. The one variation of this task that people are surprisingly good at getting right is when the rule has to do with cheating and privilege. For example, change the four cards to children in a family—"gets dessert," "doesn't get dessert," "ate vegetables," and "didn't eat vegetables"—and change the rule to "If a child gets dessert, he or she ate his or her vegetables." Many people—65 to 80 percent—get it right immediately. They turn over the "ate dessert" card, making sure the child ate his vegetables, and they turn over the "didn't eat vegetables" card, making sure the child didn't get dessert. Another way of saying this is that they turn over the "benefit received" card to make sure the cost was paid. And they turn over the "cost not paid" card to make sure no benefit was received. They look for cheaters.

The difference is startling. Subjects don't need formal logic training. They don't need math or philosophy. When asked to explain their reasoning, they say things like the answer "popped out at them."

Researchers, particularly evolutionary psychologists Leda Cosmides and John Tooby, have run this experiment with a variety of wordings and settings and on a variety of subjects: adults in the US, UK, Germany, Italy, France, and Hong Kong; Ecuadorian schoolchildren; and Shiriar tribesmen in Ecuador. The results are the same: people are bad at the Wason selection task, except when the wording involves cheating.

In the world of propositional calculus, there's absolutely no difference between a rule about traveling to Boston by plane and a rule about eating vegetables to get dessert. But in our brains, there's an enormous difference: the first is a arbitrary rule about the world, and the second is a rule of social exchange. It's of the form "If you take Benefit B, you must first satisfy Requirement R."

Our brains are optimized to detect cheaters in a social exchange. We're good at it. Even as children, we intuitively notice when someone gets a benefit he didn't pay the cost for. Those of us who grew up with a sibling have experienced how the one child not only knew that the other cheated, but felt compelled to announce it to the rest of the family. As adults, we might have learned that life isn't fair, but we still know who among our friends cheats in social exchanges. We know who doesn't pay his or her fair share of a group meal. At an airport, we might not notice the rule "If a plane is flying internationally, then it boards 15 minutes earlier than domestic flights." But we'll certainly notice who breaks the "If you board first, then you must be a first-class passenger" rule. □

***Bruce Schneier** is chief security technology officer at BT. Read his blog at www.schneier.com.*