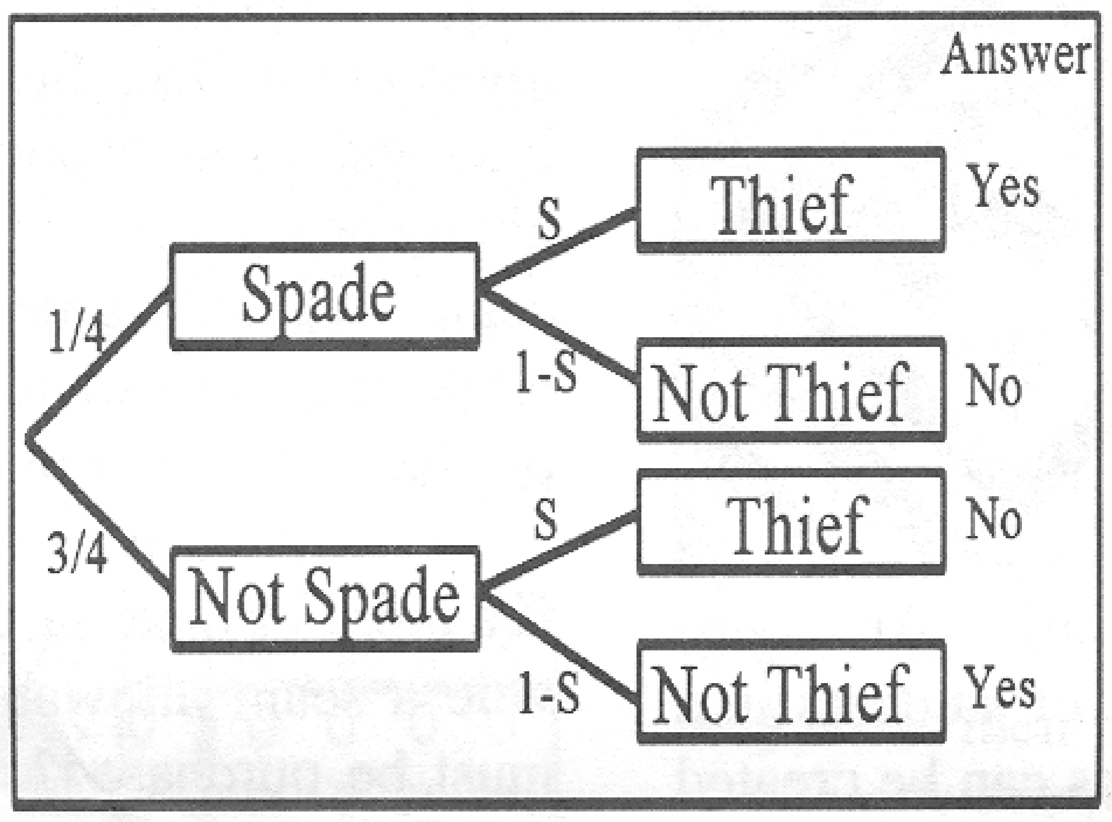
You want to know WHAT?

Paul A. Dreyer, Jr.

When pollsters ask people who they are going to vote for in an upcoming election, people usually have few qualms about telling the pollster who they support. But imagine that a pollster asked the question, "Have you ever shoplifted anything from a store?" If a person has stolen something, chances are that person will still say "no" because it is embarrassing, and no one has any reason to expose the skeletons in their closet to a pollster. Nevertheless, pollsters could still ask questions on sensitive topics and get pretty accurate results. How they could do this provides an entertaining excursion into probability.

Imagine a pollster walks up to you and says, "I'd like you pick a card from this deck. If it is a spade, I would like you to answer the question, 'Have you ever shoplifted anything from a store?' truthfully. If it is any other card, I want you to lie." Whatever you say, the pollster has no idea whether you told the truth or not, so no information about you is imparted to the pollster. Surprisingly, when many people are polled this way, it is possible to approximate what percentage of the people have actually shoplifted.



$Y = (1/4)S + (3/4)(1-S)$

Let us consider this from the pollster's point of view. When the pollster meets a person whom he wishes to poll, there are three random events that take place:

1) The person either draws a spade or does not draw a spade,
2) The person either is a shoplifter or is not a shoplifter, and
3) The person answers either "yes" or "no." The combination of these events, together with all possible outcomes, are shown in the tree diagram in the figure, where S represents the probability that a person is a shoplifter, and "Thief" is the title unflatteringly assigned to those who shoplift. For example, if a shoplifter approaches the pollster and draws a heart, he will respond "no" to the question, as indicated by the bottom "no" in the figure.

Let us use Y to denote the probability that a person who is polled answers "yes" to the question. Then we have three probabilities to consider:

- 1/4, the probability that a spade is drawn,
- S, the probability that a person is a shoplifter, and
- Y, the probability that a person says yes.

(Notice that this implies that the probability that a person draws a spade is 1–1/4, or 3/4, the probability that a person is not a shoplifter is $1–S$, and the probability that a person answers "no" to the question is $1–Y$.)

But now we make a crucial observation: A person's response *depends* on whether or not they shoplift and on the card that they draw, so we are able to write down an equation describing this relationship. First notice that there are two cases in which a person answers "yes" to the question:

- the person draws a spade and is a shoplifter, and
- the person draws a non-spade and is not a shoplifter.

Since the card a person draws is independent of their shoplifting habits, we can use the product rule to determine that "the probability that a person draws a spade and is a shoplifter" equals $(1/4)\cdot S$. Similarly, "the probability that a person draws a non-spade and is not a shoplifter" equals $(3/4)\cdot(1–S)$. Since these two cases are mutually exclusive, and since these are the only ways that we can get "yes" as an answer, we use the addition rule for probability to see that $Y = (1/4)\cdot S + (3/4)\cdot(1–S)$, as shown beneath the figure.

But we are interested in discovering S based on the number of people who answer "yes." So we solve that equation for S and obtain $S = (3–4\cdot Y)/2$.

Let us try an example: In a classroom with 30 children, suppose we ask the question "Are you in love" (a particularly prickly question!) If 10 of the students answer "yes", that that gives Y a value of 10/30 or 1/3. Thus $S = (3–4\cdot(1/3))/2$, or $S = 5/6$. So we conclude that about 5/6 of the students, that is, 25 students are in love. That answer is only approximate, of course, since this is all based on probability. Nevertheless, the outcome is striking—although only 10 out of 30 said "yes", about 25 are actually in love.

This provides a fun project for students to work on. In particular, for younger students, it is always amusing to find out what questions they find sensitive. Also, it is possible to set up more interesting randomizing systems. For example, a pollster can say, "If you pick a spade, just say yes." A warning, however. Consider the system "Flip a coin. If it is a head, lie; if a tail, tell the truth." Returning to the equation above, we get $Y = (1/2)S + (1/2)(1 – S) = 1/2$, *always*. This means that Y gives no information about S. Students can figure out what are valid and invalid randomization systems. I hope you and your students have fun with this, I certainly have!

I worked at an elementary school while going to college. I missed a month of work because I donated a kidney. When I returned, one of my kindergarten students tugged on my pants and asked, "Ms. Price, how did they get a kitty out of you?!" — Carol Price (LP '94)