**Solution**

Let G = The lot came from Guatemala. Pr(G) = 0.6.
Let H = The lot came from Honduras. Pr(H) = 0.4.
Let T = The lot had a tarantula on it. Pr(T/G) = 0.06. Pr(T/H) = 0.03.

$$\Pr(G/T) = \frac{\Pr(G)\Pr(T/G)}{\Pr(G)\Pr(T/G) + \Pr(H)\Pr(T/H)}$$

*Answer*: $\Pr(G/T) = (.6 \times .06) / [(.6 \times .06) + (.4 \times .03)] = 3/4$

### TAXICABS: ODD QUESTION 5

Here is Odd Question 5.

You have been called to jury duty in a town where there are two taxi companies, Green Cabs Ltd. and Blue Taxi Inc. Blue Taxi uses cars painted blue; Green Cabs uses green cars.

Green Cabs dominates the market, with 85% of the taxis on the road.

On a misty winter night a taxi sideswiped another car and drove off. A witness says it was a blue cab.

The witness is tested under conditions like those on the night of the accident, and 80% of the time she correctly reports the color of the cab that is seen. That is, regardless of whether she is shown a blue or a green cab in misty evening light, she gets the color right 80% of the time.

You conclude, on the basis of this information:

_____(a) The probability that the sideswiper was blue is 0.8.
_____(b) It is more likely that the sideswiper was blue, but the probability is less than 0.8.
_____(c) It is just as probable that the sideswiper was green as that it was blue.
_____(d) It is more likely than not that the sideswiper was green.

This question, like Odd Question 2, was invented by Amos Tversky and Daniel Kahneman. They have done very extensive psychological testing on this question, and found that many people think that (a) or (b) is correct. Very few think that (d) is correct. Yet (d) is, in the natural probability model, the right answer! Here is how Bayes' Rule answers the question.

**Solution**

Let G = A taxi selected at random is green. Pr(G) = 0.85.
Let B = A taxi selected at random is blue. Pr(B) = 0.15.
Let $W_b$ = The witness states that the taxi is blue.
$\Pr(W_b/B) = 0.8$.

Moreover, $\Pr(W_b/G) = 0.2$, because the witness gives a *wrong* answer 20% of the time, so the probability that she says "blue" when the cab was green is 20%.

We require $\Pr(B/W_b)$ and $\Pr(G/W_b)$.



$$\Pr(B/W_b) = \frac{\Pr(B)\Pr(W_b/B)}{\Pr(B)\Pr(W_b/B) + \Pr(G)\Pr(W_b/G)}$$

$\Pr(B/W_b) = (.15 \times .8) / [(.15 \times .8) + (.85 \times .2)] = 12/29 \approx 0.41$

*Answer*:

$\Pr(B/W_b) \approx 0.41$.
$\Pr(G/W_b) \approx 1 - 0.41 = 0.59$.
It is more likely that the sideswiper was green.

### BASE RATES

Why do so few people feel, intuitively, that (d) is the right answer? Tversky and Kahneman argue that people tend to ignore the *base rate* or background information. We focus on the fact that the witness is right 80% of the time. We ignore the fact that most of the cabs in town are green.

Suppose that we made a great many experiments with the witness, randomly selecting cabs and showing them to her on a misty night. If 100 cabs were picked at random, then we'd expect something like this:

The witness sees about 85 green cabs. She correctly identifies 80% of these as green: about 68.
She incorrectly identifies 20% as blue: about 17.
She sees about 15 blue cabs. She correctly identifies 80% of these as blue: about 12.
She incorrectly identifies 20% as green: about 3.

*So the witness identifies about 29 cabs as blue, but only 12 of these are blue!* In fact, the more we think of the problem as one about frequencies, the clearer the Bayesian answer becomes.

Some critics say that the taxicab problem does not show that we make mistakes easily. The question is asked in the wrong way. If we had been asked just about frequencies, say the critics, we would have given pretty much the right answer straightaway!

### RELIABILITY

Our witness was pretty reliable: right 80% of the time. How can a reliable witness not be trustworthy? Because of the base rates. We tend to confuse two different ideas of "reliability."

*Idea 1*: $\Pr(W_b/B)$: How reliable is she at identifying a cab as blue, given that it is in fact blue? This is a characteristic of the witness and her perceptual acumen.

*Idea 2*: $\Pr(B/W_b)$: How well can what the witness said be relied on, given that she said the cab is blue? This is a characteristic of the witness and the base rate.