

CAUSES TRUMP STATISTICS

Consider the following scenario and note your intuitive answer to the question.

A cab was involved in a hit-and-run accident at night.
Two cab companies, the Green and the Blue, operate in the city.

You are given the following data:

- 85% of the cabs in the city are Green and 15% are Blue.
- A witness identified the cab as Blue. The court tested the reliability of the witness under the circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue rather than Green?

This is a standard problem of Bayesian inference. There are two items of information: a base rate and the imperfectly reliable testimony of a witness. In the absence of a witness, the probability of the guilty cab being Blue is 15%, which is the base rate of that outcome. If the two cab companies had been equally large, the base rate would be uninformative and you would consider only the reliability of the witness, concluding that the probability is 80%. The two sources of information can be combined by Bayes's rule. The correct answer is 41%. However, you can probably guess what people do when faced with this problem: they ignore the base rate and go with the witness. The most common answer is 80%.

CAUSAL STEREOTYPES

Now consider a variation of the same story, in which only the presentation of the base rate has been altered.

You are given the following data:

- The two companies operate the same number of cabs, but Green cabs are involved in 85% of accidents.
- The information about the witness is as in the previous version.

The two versions of the problem are mathematically indistinguishable, but they are psychologically quite different. People who read the first version do not know how to use the base rate and often ignore it. In contrast, people who see the second version give considerable weight to the base rate, and their average judgment is not too far from the Bayesian solution. Why?

In the first version, the base rate of Blue cabs is a statistical fact about the cabs in the city. A mind that is hungry for causal stories finds nothing to chew on: How does the number of Green and Blue cabs in the city cause this cab driver to hit and run?

In the second version, in contrast, the drivers of Green cabs cause more than 5 times as many accidents as the Blue cabs do. The conclusion is immediate: the Green drivers must be a collection of reckless madmen! You have now formed a stereotype of Green recklessness, which you apply to unknown individual drivers in the company. The stereotype is easily fitted into a causal story, because recklessness is a causally relevant fact about individual cabdrivers. In this version, there are two causal stories that need to be combined or reconciled. The first is the hit and run, which naturally evokes the idea that a reckless Green driver was responsible. The second is the witness's testimony, which strongly suggests the cab was Blue. The inferences from the two stories about the color of the car are contradictory and approximately cancel each other. The chances for the two colors are about equal (the Bayesian estimate is 41%, reflecting the fact that the base rate of Green cabs is a little

more extreme than the reliability of the witness who reported a Blue cab).

The cab example illustrates two types of base rates. *Statistical base rates* are facts about a population to which a case belongs, but they are not relevant to the individual case. *Causal base rates* change your view of how the individual case came to be. The two types of base-rate information are treated differently:

- Statistical base rates are generally underweighted, and sometimes neglected altogether, when specific information about the case at hand is available.
- Causal base rates are treated as information about the individual case and are easily combined with other case-specific information.

The causal version of the cab problem had the form of a stereotype: Green drivers are dangerous. Stereotypes are statements about the group that are (at least tentatively) accepted as facts about every member. Here are two examples:

Most of the graduates of this inner-city school go to college.
Interest in cycling is widespread in France.

These statements are readily interpreted as setting up a propensity in individual members of the group, and they fit in a causal story. Many graduates of this particular inner-city school are eager and able to go to college, presumably because of some beneficial features of life in that school. There are forces in French culture and social life that cause many Frenchmen to take an interest in cycling. You will be reminded of these facts when you think about the likelihood that a particular graduate of the school will attend college, or when you wonder whether to bring up the Tour de France in a conversation with a Frenchman you just met.

Stereotyping is a bad word in our culture, but in my usage it is neutral. One of the basic characteristics of System 1 is that it