16

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 represents categories as norms and prototypical exemplars. This is how we think of horses, refrigerators, and New York police officers; we hold in memory a representation of one or more "normal" members of each of these categories. When the categories are social, these representations are called stereotypes. Some stereotypes are perniciously wrong, and hostile stereotyping can have dreadful

consequences, but the psychological facts cannot be avoided: stereotypes, both correct and false, are how we think of categories.

You may note the irony. In the context of the cab problem, the neglect of base-rate information is a cognitive flaw, a failure of Bayesian reasoning, and the reliance on causal base rates is desirable. Stereotyping the Green drivers improves the accuracy of judgment. In other contexts, however, such as hiring or profiling, there is a strong social norm against stereotyping, which is also embedded in the law. This is as it should be. In sensitive social contexts, we do not want to draw possibly erroneous conclusions about the individual from the statistics of the group. We consider it morally desirable for base rates to be treated as statistical facts about the group rather than as presumptive facts about individuals. In other words, we reject causal base rates.

The social norm against stereotyping, including the opposition to profiling, has been highly beneficial in creating a more civilized and more equal society. It is useful to remember, however, that neglecting valid stereotypes inevitably results in suboptimal judgments. Resistance to stereotyping is a laudable moral position, but the simplistic idea that the resistance is costless is wrong. The costs are worth paying to achieve a better society, but denying that the costs exist, while satisfying to the soul and politically correct, is not scientifically defensible. Reliance on the affect heuristic is common in politically charged arguments. The positions we favor have no cost and those we oppose have no benefits. We should be able to do better.

CAUSAL SITUATIONS

Amos and I constructed the variants of the cab problem, but we did not invent the powerful notion of causal base rates; we borrowed it from the psychologist Icek Ajzen. In his experiment, Ajzen showed his participants brief vignettes describing some students who had taken an exam at Yale and asked the participants to judge the probability that each student had passed the test. The manipulation of causal base rates was straightforward: Ajzen told one group that the students they saw had been drawn from a class in which 75% passed the exam, and told another group that the same students had been in a class in which only 25% passed. This is a powerful manipulation, because the base rate of passing suggests the immediate inference that the test that only 25% passed must have been brutally difficult. The difficulty of a test is, of course, one of the causal factors that determine every

student's outcome. As expected, Ajzen's subjects were highly sensitive to the causal base rates, and every student was judged more likely to pass in the high-success condition than in the high-failure rate.

Ajzen used an ingenious method to suggest a noncausal base rate. He told his subjects that the students they saw had been drawn from a sample, which itself was constructed by selecting students who had passed or failed the exam. For example, the information for the high-failure group read as follows:

> The investigator was mainly interested in the causes of failure and constructed a sample in which 75% had failed the examination.

Note the difference. This base rate is a purely statistical fact about the ensemble from which cases have been drawn. It has no bearing on the question asked, which is whether the individual student passed or failed the test. As expected, the explicitly stated base rates had some effects on judgment, but they had much less impact than the statistically equivalent causal base rates. System 1 can deal with stories in which the elements are causally linked, but it is weak in statistical reasoning. For a Bayesian thinker, of course, the versions are equivalent. It is tempting to conclude that we have reached a satisfactory conclusion: causal base rates are used; merely statistical facts are (more or less) neglected. The next study, one of my all-time favorites, shows that the situation is rather more complex.

CAN PSYCHOLOGY BE TAUGHT?

The reckless cabdrivers and the impossibly difficult exam illustrate two inferences that people can draw from causal base rates: a stereotypical trait that is attributed to an individual, and a significant feature of the situation that affects an individual's outcome. The participants in the experiments made the correct inferences and their judgments improved. Unfortunately, things do not always work out so well. The classic experiment I describe next shows that people will not draw from base-rate information an inference that conflicts with other beliefs. It also supports the uncomfortable conclusion that teaching psychology is mostly a waste of time.

The experiment was conducted a long time ago by the social psychologist Richard Nisbett and his student Eugene Borgida, at the University of Michigan. They told students about the renowned "helping experiment" that had been conducted a few years earlier at New York University. Partici-

pants in that experiment were led to individual booths and invited to speak over the intercom about their personal lives and problems. They were to talk in turn for about two minutes. Only one microphone was active at any one time. There were six participants in each group, one of whom was a stooge. The stooge spoke first, following a script prepared by the experimenters. He described his problems adjusting to New York and admitted with obvious embarrassment that he was prone to seizures, especially when stressed. All the participants then had a turn. When the microphone was again turned over to the stooge, he became agitated and incoherent, said he felt a seizure coming on, and asked for someone to help him. The last words heard from him were, "C-could somebody-er-er-help-er-uh-uh-uh [choking sounds]. I . . . I'm gonna die-er-er-er I'm . . . gonna die-er-er-I seizure I-er [chokes, then quiet]." At this point the microphone of the next participant automatically became active, and nothing more was heard from the possibly dying individual.

What do you think the participants in the experiment did? So far as the participants knew, one of them was having a seizure and had asked for help. However, there were several other people who could possibly respond, so perhaps one could stay safely in one's booth. These were the results: only four of the fifteen participants responded immediately to the appeal for help. Six never got out of their booth, and five others came out only well after the "seizure victim" apparently choked. The experiment shows that individuals feel relieved of responsibility when they know that others have heard the same request for help.

Did the results surprise you? Very probably. Most of us think of ourselves as decent people who would rush to help in such a situation, and we expect other decent people to do the same. The point of the experiment, of course, was to show that this expectation is wrong. Even normal, decent people do not rush to help when they expect others to take on the unpleasantness of dealing with a seizure. And that means you, too.

Are you willing to endorse the following statement? "When I read the procedure of the helping experiment I thought I would come to the stranger's help immediately, as I probably would if I found myself alone with a seizure victim. I was probably wrong. If I find myself in a situation in which other people have an opportunity to help, I might not step forward. The presence of others would reduce my sense of personal responsibility more than I initially thought." This is what a teacher of psychology would hope you would learn. Would you have made the same inferences by yourself?

The psychology professor who describes the helping experiment wants

the students to view the low base rate as causal, just as in the case of the fictitious Yale exam. He wants them to infer, in both cases, that a surprisingly high rate of failure implies a very difficult test. The lesson students are meant to take away is that some potent feature of the situation, such as the diffusion of responsibility, induces normal and decent people such as them to behave in a surprisingly unhelpful way.

Changing one's mind about human nature is hard work, and changing one's mind for the worse about oneself is even harder. Nisbett and Borgida suspected that students would resist the work and the unpleasantness. Of course, the students would be able and willing to recite the details of the helping experiment on a test, and would even repeat the "official" interpretation in terms of diffusion of responsibility. But did their beliefs about human nature really change? To find out, Nisbett and Borgida showed them videos of brief interviews allegedly conducted with two people who had participated in the New York study. The interviews were short and bland. The interviewees appeared to be nice, normal, decent people. They described their hobbies, their spare-time activities, and their plans for the future, which were entirely conventional. After watching the video of an interview, the students guessed how quickly that particular person had come to the aid of the stricken stranger.

To apply Bayesian reasoning to the task the students were assigned, you should first ask yourself what you would have guessed about the two individuals if you had not seen their interviews. This question is answered by consulting the base rate. We have been told that only 4 of the 15 participants in the experiment rushed to help after the first request. The probability that an unidentified participant had been immediately helpful is therefore 27%. Thus your prior belief about any unspecified participant should be that he did not rush to help. Next, Bayesian logic requires you to adjust your judgment in light of any relevant information about the individual. However, the videos were carefully designed to be uninformative; they provided no reason to suspect that the individuals would be either more or less helpful than a randomly chosen student. In the absence of useful new information, the Bayesian solution is to stay with the base rates.

Nisbett and Borgida asked two groups of students to watch the videos and predict the behavior of the two individuals. The students in the first group were told only about the procedure of the helping experiment, not about its results. Their predictions reflected their views of human nature

and their understanding of the situation. As you might expect, they predicted that both individuals would immediately rush to the victim's aid. The second group of students knew both the procedure of the experiment and its results. The comparison of the predictions of the two groups provides an answer to a significant question: Did students learn from the results of the helping experiment anything that significantly changed their way of thinking? The answer is straightforward: they learned nothing at all. Their predictions about the two individuals were indistinguishable from the predictions made by students who had not been exposed to the statistical results of the experiment. They knew the base rate in the group from which the individuals had been drawn, but they remained convinced that the people they saw on the video had been quick to help the stricken stranger.

For teachers of psychology, the implications of this study are disheartening. When we teach our students about the behavior of people in the helping experiment, we expect them to learn something they had not known before; we wish to change how they think about people's behavior in a particular situation. This goal was not accomplished in the Nisbett-Borgida study, and there is no reason to believe that the results would have been different if they had chosen another surprising psychological experiment. Indeed, Nisbett and Borgida reported similar findings in teaching another study, in which mild social pressure caused people to accept much more painful electric shocks than most of us (and them) would have expected. Students who do not develop a new appreciation for the power of social setting have learned nothing of value from the experiment. The predictions they make about random strangers, or about their own behavior, indicate that they have not changed their view of how they would have behaved. In the words of Nisbett and Borgida, students "quietly exempt themselves" (and their friends and acquaintances) from the conclusions of experiments that surprise them. Teachers of psychology should not despair, however, because Nisbett and Borgida report a way to make their students appreciate the point of the helping experiment. They took a new group of students and taught them the procedure of the experiment but did not tell them the group results. They showed the two videos and simply told their students that the two individuals they had just seen had not helped the stranger, then asked them to guess the global results. The outcome was dramatic: the students' guesses were extremely accurate.

To teach students any psychology they did not know before, you must surprise them. But which surprise will do? Nisbett and Borgida found that when they presented their students with a surprising statistical fact, the

students managed to learn nothing at all. But when the students were surprised by individual cases—two nice people who had not helped—they immediately made the generalization and inferred that helping is more difficult than they had thought. Nisbett and Borgida summarize the results in a memorable sentence:

> Subjects' unwillingness to deduce the particular from the general was matched only by their willingness to infer the general from the particular.

This is a profoundly important conclusion. People who are taught surprising statistical facts about human behavior may be impressed to the point of telling their friends about what they have heard, but this does not mean that their understanding of the world has really changed. The test of learning psychology is whether your understanding of situations you encounter has changed, not whether you have learned a new fact. There is a deep gap between our thinking about statistics and our thinking about individual cases. Statistical results with a causal interpretation have a stronger effect on our thinking than noncausal information. But even compelling causal statistics will not change long-held beliefs or beliefs rooted in personal experience. On the other hand, surprising individual cases have a powerful impact and are a more effective tool for teaching psychology because the incongruity must be resolved and embedded in a causal story. That is why this book contains questions that are addressed personally to the reader. You are more likely to learn something by finding surprises in your own behavior than by hearing surprising facts about people in general.

SPEAKING OF CAUSES AND STATISTICS

"We can't assume that they will really learn anything from mere statistics. Let's show them one or two representative individual cases to influence their System 1."

"No need to worry about this statistical information being ignored. On the contrary, it will immediately be used to feed a stereotype."

17

REGRESSION TO THE MEAN

I had one of the most satisfying eureka experiences of my career whil teaching flight instructors in the Israeli Air Force about the psychology c effective training. I was telling them about an important principle of ski training: rewards for improved performance work better than punishmen of mistakes. This proposition is supported by much evidence from researc on pigeons, rats, humans, and other animals.

When I finished my enthusiastic speech, one of the most seasoned in structors in the group raised his hand and made a short speech of his ow He began by conceding that rewarding improved performance might b good for the birds, but he denied that it was optimal for flight cadets. Thi is what he said: "On many occasions I have praised flight cadets for clea execution of some aerobatic maneuver. The next time they try the same ma neuver they usually do worse. On the other hand, I have often screame into a cadet's earphone for bad execution, and in general he does better o his next try. So please don't tell us that reward works and punishment doe not, because the opposite is the case."

This was a joyous moment of insight, when I saw in a new light a prin ciple of statistics that I had been teaching for years. The instructor wa right—but he was also completely wrong! His observation was astute an correct: occasions on which he praised a performance were likely to be fol lowed by a disappointing performance, and punishments were typicall followed by an improvement. But the inference he had drawn about the ef ficacy of reward and punishment was completely off the mark. What he ha