Tom W's Specialty

Have a look at a simple puzzle:

> Tom W is a graduate student at the main university in your state. Please rank the following nine fields of graduate specialization in order of the likelihood that Tom W is now a student in each of these fields. Use 1 for the most likely, 9 for the least likely.
>
> business administration
> computer science
> engineering
> humanities and education
> law
> medicine
> library science
> physical and life sciences
> social science and social work

This question is easy, and you knew immediately that the relative size of enrollment in the different fields is the key to a solution. So far as you know, Tom W was picked at random from the graduate students at the university, like a single marble drawn from an urn. To decide whether a marble is more likely to be red or green, you need to know how many marbles of each color there are in the urn. The proportion of marbles of a particular kind is called a *base rate*. Similarly, the base rate of humanities and education in this problem is the proportion of students of that field among all the graduate students. In the absence of specific information about Tom W, you will go by the base rates and guess that he is more likely to be enrolled in humanities and education than in computer science or library science, because there are more students overall in the humanities and education than in the other two fields. Using base-rate information is the obvious move when no other information is provided.

Next comes a task that has nothing to do with base rates.

> The following is a personality sketch of Tom W written during Tom's senior year in high school by a psychologist, on the basis of psychological tests of uncertain validity:

> Tom W is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people, and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense.
>
> Now please take a sheet of paper and rank the nine fields of specialization listed below by how similar the description of Tom W is to the typical graduate student in each of the following fields. Use 1 for the most likely and 9 for the least likely.

You will get more out of the chapter if you give the task a quick try; reading the report on Tom W is necessary to make your judgments about the various graduate specialties.

This question too is straightforward. It requires you to retrieve, or perhaps to construct, a stereotype of graduate students in the different fields. When the experiment was first conducted, in the early 1970s, the average ordering was as follows. Yours is probably not very different:

1. computer science
2. engineering
3. business administration
4. physical and life sciences
5. library science
6. law
7. medicine
8. humanities and education
9. social science and social work

You probably ranked computer science among the best fitting because of hints of nerdiness (“corny puns”). In fact, the description of Tom W was written to fit that stereotype. Another specialty that most people ranked high is engineering (“neat and tidy systems”). You probably thought that Tom W is not a good fit with your idea of social science and social work

("little feel and little sympathy for other people"). Professional stereotypes appear to have changed little in the nearly forty years since I designed the description of Tom W.

The task of ranking the nine careers is complex and certainly requires the discipline and sequential organization of which only System 2 is capable. However, the hints planted in the description (corny puns and others) were intended to activate an association with a stereotype, an automatic activity of System 1.

The instructions for this similarity task required a comparison of the description of Tom W to the stereotypes of the various fields of specialization. For the purposes of tv>

If you examine Tom W again, you will see that he is a good fit to stereotypes of some small groups of students (computer scientists, librarians, engineers) and a much poorer fit to the largest groups (humanities and education, social science and social work). Indeed, the participants almost always ranked the two largest fields very low. Tom W was intentionally designed as an "anti-base-rate" character, a good fit to small fields and a poor fit to the most populated specialties.

Predicting by Representativeness

The third task in the sequence was administered to graduate students in psychology, and it is the critical one: rank the fields of specialization in order of the likelihood that Tom W is now a graduate student in each of these fields. The members of this prediction group knew the relevant statistical facts: they were familiar with the base rates of the different fields, and they knew that the source of Tom W's description was not highly trustworthy. However, we expected them to focus exclusively on the similarity of the description to the stereotypes—we called it *representativeness*—ignoring both the base rates and the doubts about the veracity of the description. They would then rank the small specialty—computer science—as highly probable, because that outcome gets the highest representativeness score.

Amos and I worked hard during the year we spent in Eugene, and I sometimes stayed in the office through the night. One of my tasks for such a night was to make up a description that would pit representativeness and base rates against each other. Tom W was the result of my efforts, and I completed the description in the early morning hours. The first person who showed up to work that morning was our colleague and friend Robyn Dawes, who was both a sophisticated statistician and a skeptic about the validity of intuitive judgment. If anyone would see the relevance of the base

rate, it would have to be Robyn. I called Robyn over, gave him the question I had just typed, and asked him to guess Tom W's profession. I still remember his sly smile as he said tentatively, "computer scientist?" That was a happy moment—even the mighty had fallen. Of course, Robyn immediately recognized his mistake as soon as I mentioned "base rate," but he had not spontaneously thought of it. Although he knew as much as anyone about the role of base rates in prediction, he neglected them when presented with the description of an individual's personality. As expected, he substituted a judgment of representativeness for the probability he was asked to assess.

Amos and I then collected answers to the same question from 114 graduate students in psychology at three major universities, all of whom had taken several courses in statistics. They did not disappoint us. Their rankings of the nine fields by probability did not differ from ratings by similarity to the stereotype. Substitution was perfect in this case: there was no indication that the participants did anything else but judge representativeness. The question about probability (likelihood) was difficult, but the question about similarity was easier, and it was answered instead. This is a serious mistake, because judgments of similarity and probak tbility are not constrained by the same logical rules. It is entirely acceptable for judgments of similarity to be unaffected by base rates and also by the possibility that the description was inaccurate, but anyone who ignores base rates and the quality of evidence in probability assessments will certainly make mistakes.

The concept "the probability that Tom W studies computer science" is not a simple one. Logicians and statisticians disagree about its meaning, and some would say it has no meaning at all. For many experts it is a measure of subjective degree of belief. There are some events you are sure of, for example, that the sun rose this morning, and others you consider impossible, such as the Pacific Ocean freezing all at once. Then there are many events, such as your next-door neighbor being a computer scientist, to which you assign an intermediate degree of belief—which is your probability of that event.

Logicians and statisticians have developed competing definitions of probability, all very precise. For laypeople, however, probability (a synonym of *likelihood* in everyday language) is a vague notion, related to uncertainty, propensity, plausibility, and surprise. The vagueness is not particular to this concept, nor is it especially troublesome. We know, more or less, what we mean when we use a word such as *democracy* or *beauty* and the people we are talking to understand, more or less, what we intended to say. In all the years I spent asking questions about the

probability of events, no one ever raised a hand to ask me, “Sir, what do you mean by probability?” as they would have done if I had asked them to assess a strange concept such as globability. Everyone acted as if they knew how to answer my questions, although we all understood that it would be unfair to ask them for an explanation of what the word means.

People who are asked to assess probability are not stumped, because they do not try to judge probability as statisticians and philosophers use the word. A question about probability or likelihood activates a mental shotgun, evoking answers to easier questions. One of the easy answers is an automatic assessment of representativeness—routine in understanding language. The (false) statement that “Elvis Presley’s parents wanted him to be a dentist” is mildly funny because the discrepancy between the images of Presley and a dentist is detected automatically. System 1 generates an impression of similarity without intending to do so. The representativeness heuristic is involved when someone says “She will win the election; you can see she is a winner” or “He won’t go far as an academic; too many tattoos.” We rely on representativeness when we judge the potential leadership of a candidate for office by the shape of his chin or the forcefulness of his speeches.

Although it is common, prediction by representativeness is not statistically optimal. Michael Lewis’s bestselling *Moneyball* is a story about the inefficiency of this mode of prediction. Professional baseball scouts traditionally forecast the success of possible players in part by their build and look. The hero of Lewis’s book is Billy Beane, the manager of the Oakland A’s, who made the unpopular decision to overrule his scouts and to select players by the statistics of past performance. The players the A’s picked were inexpensive, because other teams had rejected them for not looking the part. The team soon achieved excellent results at low cost.

The Sins of Representativeness

Judging probability byals representativeness has important virtues: the intuitive impressions that it produces are often—indeed, usually—more accurate than chance guesses would be.

- On most occasions, people who act friendly are in fact friendly.
- A professional athlete who is very tall and thin is much more likely to play basketball than football.
- People with a PhD are more likely to subscribe to *The New York Times* than people who ended their education after high school.

- Young men are more likely than elderly women to drive aggressively.

In all these cases and in many others, there is some truth to the stereotypes that govern judgments of representativeness, and predictions that follow this heuristic may be accurate. In other situations, the stereotypes are false and the representativeness heuristic will mislead, especially if it causes people to neglect base-rate information that points in another direction. Even when the heuristic has some validity, exclusive reliance on it is associated with grave sins against statistical logic.

One sin of representativeness is an excessive willingness to predict the occurrence of unlikely (low base-rate) events. Here is an example: you see a person reading *The New York Times* on the New York subway. Which of the following is a better bet about the reading stranger?

She has a PhD.
She does not have a college degree.

Representativeness would tell you to bet on the PhD, but this is not necessarily wise. You should seriously consider the second alternative, because many more nongraduates than PhDs ride in New York subways. And if you must guess whether a woman who is described as "a shy poetry lover" studies Chinese literature or business administration, you should opt for the latter option. Even if every female student of Chinese literature is shy and loves poetry, it is almost certain that there are more bashful poetry lovers in the much larger population of business students.

People without training in statistics are quite capable of using base rates in predictions under some conditions. In the first version of the Tom W problem, which provides no details about him, it is obvious to everyone that the probability of Tom W's being in a particular field is simply the base rate frequency of enrollment in that field. However, concern for base rates evidently disappears as soon as Tom W's personality is described.

Amos and I originally believed, on the basis of our early evidence, that base-rate information will *always* be neglected when information about the specific instance is available, but that conclusion was too strong. Psychologists have conducted many experiments in which base-rate information is explicitly provided as part of the problem, and many of the participants are influenced by those base rates, although the information about the individual case is almost always weighted more than mere statistics. Norbert Schwarz and his colleagues showed that instructing people to "think like a statistician" enhanced the use of base-rate information, while the instruction to "think like a clinician" had the opposite

effect.

An experiment that was conducted a few years ago with Harvard undergradut oates yielded a finding that surprised me: enhanced activation of System 2 caused a significant improvement of predictive accuracy in the Tom W problem. The experiment combined the old problem with a modern variation of cognitive fluency. Half the students were told to puff out their cheeks during the task, while the others were told to frown. Frowning, as we have seen, generally increases the vigilance of System 2 and reduces both overconfidence and the reliance on intuition. The students who puffed out their cheeks (an emotionally neutral expression) replicated the original results: they relied exclusively on representativeness and ignored the base rates. As the authors had predicted, however, the frowners did show some sensitivity to the base rates. This is an instructive finding.

When an incorrect intuitive judgment is made, System 1 and System 2 should both be indicted. System 1 suggested the incorrect intuition, and System 2 endorsed it and expressed it in a judgment. However, there are two possible reasons for the failure of System 2—ignorance or laziness. Some people ignore base rates because they believe them to be irrelevant in the presence of individual information. Others make the same mistake because they are not focused on the task. If frowning makes a difference, laziness seems to be the proper explanation of base-rate neglect, at least among Harvard undergrads. Their System 2 “knows” that base rates are relevant even when they are not explicitly mentioned, but applies that knowledge only when it invests special effort in the task.

The second sin of representativeness is insensitivity to the quality of evidence. Recall the rule of System 1: WYSIATI. In the Tom W example, what activates your associative machinery is a description of Tom, which may or may not be an accurate portrayal. The statement that Tom W “has little feel and little sympathy for people” was probably enough to convince you (and most other readers) that he is very unlikely to be a student of social science or social work. But you were explicitly told that the description should not be trusted!

You surely understand in principle that worthless information should not be treated differently from a complete lack of information, but WY SIATI makes it very difficult to apply that principle. Unless you decide immediately to reject evidence (for example, by determining that you received it from a liar), your System 1 will automatically process the information available as if it were true. There is one thing you can do when you have doubts about the quality of the evidence: let your judgments of

probability stay close to the base rate. Don't expect this exercise of discipline to be easy—it requires a significant effort of self-monitoring and self-control.

The correct answer to the Tom W puzzle is that you should stay very close to your prior beliefs, slightly reducing the initially high probabilities of well-populated fields (humanities and education; social science and social work) and slightly raising the low probabilities of rare specialties (library science, computer science). You are not exactly where you would be if you had known nothing at all about Tom W, but the little evidence you have is not trustworthy, so the base rates should dominate your estimates.

How to Discipline Intuition

Your probability that it will rain tomorrow is your subjective degree of belief, but you should not let yourself believe whatever comes to your mind. To be useful, your beliefs should be constrained by the logic of probability. So if you believe that there is a 40% chance plethat it will rain sometime tomorrow, you must also believe that there is a 60% chance it will not rain tomorrow, and you must not believe that there is a 50% chance that it will rain tomorrow morning. And if you believe that there is a 30% chance that candidate X will be elected president, and an 80% chance that he will be reelected if he wins the first time, then you must believe that the chances that he will be elected twice in a row are 24%.

The relevant "rules" for cases such as the Tom W problem are provided by Bayesian statistics. This influential modern approach to statistics is named after an English minister of the eighteenth century, the Reverend Thomas Bayes, who is credited with the first major contribution to a large problem: the logic of how people should change their mind in the light of evidence. Bayes's rule specifies how prior beliefs (in the examples of this chapter, base rates) should be combined with the diagnosticity of the evidence, the degree to which it favors the hypothesis over the alternative. For example, if you believe that 3% of graduate students are enrolled in computer science (the base rate), and you also believe that the description of Tom W is 4 times more likely for a graduate student in that field than in other fields, then Bayes's rule says you must believe that the probability that Tom W is a computer scientist is now 11%. If the base rate had been 80%, the new degree of belief would be 94.1%. And so on.

The mathematical details are not relevant in this book. There are two ideas to keep in mind about Bayesian reasoning and how we tend to mess it up. The first is that base rates matter, even in the presence of evidence about the case at hand. This is often not intuitively obvious. The second is

that intuitive impressions of the diagnosticity of evidence are often exaggerated. The combination of WY SIATI and associative coherence tends to make us believe in the stories we spin for ourselves. The essential keys to disciplined Bayesian reasoning can be simply summarized:

- Anchor your judgment of the probability of an outcome on a plausible base rate.
- Question the diagnosticity of your evidence.

Both ideas are straightforward. It came as a shock to me when I realized that I was never taught how to implement them, and that even now I find it unnatural to do so.

Speaking of Representativeness

> "The lawn is well trimmed, the receptionist looks competent, and the furniture is attractive, but this doesn't mean it is a well-managed company. I hope the board does not go by representativeness."

> "This start-up looks as if it could not fail, but the base rate of success in the industry is extremely low. How do we know this case is different?"

> "They keep making the same mistake: predicting rare events from weak evidence. When the evidence is weak, one should stick with the base rates."

> "I know this report is absolutely damning, and it may be based on solid evidence, but how sure are we? We must allow for that uncertainty in our thinking."

ht="5%">