3.7: The Base Rate Fallacy

Consider the following scenario. You go in for some testing for some health problems you’ve been having and after a number of tests, you test positive for colon cancer. What are the chances that you really do have colon cancer? Let’s suppose that the test is not perfect, but it is 95% accurate. That is, in the case of those who really do have colon cancer, the test will detect the cancer 95% of the time (and thus miss it 5% of the time). (The test will also misdiagnose those who don’t actually have colon cancer 5% of the time.) Many people would be inclined to say that, given the test and its accuracy, there is a 95% chance that you have colon cancer. However, if you are like most people and are inclined to answer this way, you are wrong. In fact, you have committed the fallacy of ignoring the base rate (i.e., the base rate fallacy).

The base rate in this example is the rate of those who have colon cancer in a population. There is very small percentage of the population that actually has colon cancer (let’s suppose it is .005 or .5%), so the probability that you have it must take into account the very low probability that you are one of the few that have it. That is, prior to the test (and not taking into account any other details about you), there was a very low probability that you have it—that is, a half of one percent chance (.5%). The test is 95% accurate, but given the very low prior probability that you have colon cancer, we cannot simply now say that there is a 95% chance that you have it. Rather, we must temper that figure with the very low base rate. Here is how we do it. Let’s suppose that our population is 100,000 people. If we were to apply the test to that whole population, it would deliver 5000 false positives. A false positive occurs when a test registers that some feature is present, when the feature isn’t really present. In this case, the false positive is when the test for colon cancer (which will give false positives in 5% of the cases) says that someone has it when they really don’t. The number of people who actually have colon cancer (based on the stated base rate) is 500, and the test will accurately identify 95 percent of those (or 475 people). So what you need to know is the probability that you are one who tested positive and actually has colon cancer rather than one of the false positives. And what is the probability of that? It is simply the number of people who actually have colon cancer (500) divided by the number that the test would identify as having colon cancer. This latter number includes those the test would misidentify (5000) as well as the number it would accurately identify.
Thus the total number the test would identify as having colon cancer would be 5475. So the probability that you have it, given the positive test = 500/5475 = .091 or 9.1%. So the probability that you have cancer, given the evidence of the positive test is 9.1%. Thus, contrary to our initial reasoning that there was a 95% chance that you have colon cancer, the chance is only a tenth of that—it is less than 10%! In thinking that the probability that you have cancer is closer to 95% you would be ignoring the base rate of the probability of having the disease in the first place (which, as we’ve seen, is quite low). This is the signature of any base rate fallacy. Before closing this section, let’s look at one more example of a base rate fallacy.

Suppose that the government has developed a machine that is able to detect terrorist intent with an accuracy of 90%. During a joint meeting of congress, a highly trustworthy source says that there is a terrorist in the building. (Let’s suppose, for the sake of simplifying this example, that there is in fact a terrorist in the building.) In order to determine who the terrorist is, the building security seals all the exits, rounds up all 3000 people in the building and uses the machine to test each person. The first 30 people pass without triggering a positive identification from the machine, but on the very next person, the machine triggers a positive identification of terrorist intent. The question is: what are the chances that the person who set off the machine really is a terrorist? Consider the following three possibilities: a) 90%, b) 10%, or c) .3%. If you answered 90%, then you committed the base rate fallacy again. The actually answer is “c” less than 1%! Here is the relevant reasoning. The base rate here is that it is exceedingly unlikely that any individual is a terrorist, given that there is only one terrorist in the building and there are 3000 people in the building. That means the probability of any one person being a terrorist, before any results of the test, is exceedingly low: 1/3000. Since the test is 90% accurate, that means that out of the 3000 people, it will misidentify 10% of them as terrorists = 300 false positives. Assuming the machine doesn’t misidentify the one actual terrorist, the machine will identify a total of 301 individuals as “possessing terrorist intent.” The probability that any one of them actually This is another good illustration of how far off probabilities can be when the base rate is ignored.

This example is taken (with certain alterations) from: http://news.bbc.co.uk/2/hi/uk_news/8153539.stm