Epistemology and the Psychology of Human Judgment
by Michael A. Bishop and J.D. Trout

Chapter 8 Putting Epistemology into Practice
Human irrationality

- The question posed in this chapter is whether it is possible to give evidence of human irrationality. This is an odd question, since most of us would see that we see evidence of this all the time, but you have done enough philosophy at this point to realize that any claim to the contrary probably has something to do with the is-ought distinction. Specifically, since rationality is an inherently normative concept we should expect some kind of argument that it is impossible to give strictly empirical evidence of human irrationality.

- It may also help to recall Plato’s argument that humans, by their very nature, always seek the good. Plato believed that it was actually impossible to desire evil for its own sake. Even people who intentionally cause great suffering in others believe themselves to be doing good.

- Plato argued that all evil is the result of ignorance, and this is the kind of argument that some contemporary philosophers make about human rationality as well: all people are fundamentally rational. Apparently irrational behavior is the result of ignorance, either on the part of the agent or the part of the individual who is judging the behavior irrational.
Heuristics and biases

- During the past twenty-five years researchers like Tversky and Kahneman have provided a great deal of empirical evidence that seems most plausibly described as evidence that humans are actually quite illogical. There are many different examples of intelligent people who fail at reasoning tasks that one would expect minimally proficient reasoners to perform well at.

- One of these, developed in the early 1960’s is called the Wason Selection task. In this experiment cards are used in which the subjects know that one side of the card has a number and the other has a letter. Four cards are placed faced down like so and asked the following question.

- The vast majority of people do not get this right.

- The general question here is whether we can conclude from this that people are poor at reasoning or whether these results can be explained away somehow.

- The generally favored view today is that human beings navigate through the world using rules of thumb (heuristics) and biases that allow them to make quick and dirty calculations that actually work in many cases but fail quite spectacularly in others.

Each of the cards above has a number on one side and a letter on the other.

Question: Which cards must you turn over in order to know whether the following statement is true:

*If there is an R on one side, then there is a 2 on the other.*
Reject the norm

- B&T consider two conceptual arguments for the conclusion that empirical results like these do not, and in fact can not, give evidence of human irrationality.
- They call these arguments “reject-the-norm” arguments because they both involve the claim that researchers are using improper norms for measuring the competence of human reasoners.
- B&T are careful to distinguish conceptual reject-the-norm arguments from empirical reject-the-norm arguments. Empirical arguments involve the claim that subjects make what appear to be reasoning mistakes simply because they misunderstood the problem. These things surely do happen, but they can be dealt with by improving the experimental design.
- Conceptual reject-the-norm arguments are *a priori* in that they do not rely on any empirical facts about the subjects.
One of these conceptual arguments focuses on the concept of base rate neglect, so it will be helpful to review this concept.

The base rate is essentially the prior probability of something occurring. For example, if you are being tested for some medical condition C, it is important to know the base rate of that condition, i.e., its actual frequency in the population, in order to determine how to evaluate a positive or negative test result.

For example, suppose you are being tested for a particularly nasty condition known as philosophitis, which typically results in brain death in about 4 years. Philosophitis is extremely rare. It occurs in only about .5% of the population.

Fortunately, there is a simple test for philosophitis which is 99% accurate. This means that if you have philosophitis the test will reveal it about 99% of the time.

All tests have error rates. Sometimes they say you don’t have the condition when you do. (This is called a false negative.) Other times they say you do have the condition when you don’t. (This is called a false positive.)

From the 99% accuracy rate above we can deduce a false negative rate of 1%. You can’t derive the false positive rate from the information provided so far. The false positive rate for this particular test is also 1%.

So, now that you’ve tested positive for philosophitis, the question is: What is the likelihood that you actually have it?
There is a mathematical formula for figuring this out precisely. It is called Bayes Rule, and it can be derived from some very simple facts about probability. We won’t do the derivations, but it’s important to understand what the rule says.

In words, Bayes rule says that the probability that you have condition C given that you have received a positive test T is a certain quotient.

The dividend of the quotient says: the probability that you test positive given that you have the condition multiplied by the probability that you have the condition.

The divisor of the quotient is the sum of the dividend and the probability that you will test positive given that you do not have the condition multiplied by the probability of the condition.

For our example of philosophitis, Bayes rules says that you can figure out the probability that you have philosophitis given that you have tested positive by multiplying the effectiveness of the test by the base rate (.99 x .005). Then you divide this by itself plus what you get when you multiply the false positive rate x the base rate of the condition’s non occurrence. (.01 x .9995).

Bayes Rule:

\[
P(C/T) = \frac{P(T/C) \times P(C)}{[P(T/C) \times P(C)] + [P(T/-C) \times P(-C)]}
\]

\[
P(C/T) = \frac{.99 \times .005}{(.99 \times .005) + (.01 \times .9995)}
\]

\[
P(C/T) = 33%
\]
The important point to understand is that the unreasonably high probabilities that people attach to the results of the test are due, at least partially, to neglecting the effect of the base rate.

The basic idea is that when the base rate is very low, then the false positives actually tend to account for a far greater percentage of positive tests than the true positives.

You do not need Bayes Rule to figure this out. A different approach to the same problem is just to draw a simple probability matrix. For our philosophitis case assume a population of 20,000 people. With a .005 base rate, this means 100 people will actually have philosophitis.

The first column tells the story. Of the 298 people who will test positive, only 99 of them actually have philosophitis.

<table>
<thead>
<tr>
<th></th>
<th>Tests positive for P</th>
<th>Tests negative for P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has P</td>
<td>.99 x 100 = 99</td>
<td>.01 x 100 = 1</td>
</tr>
<tr>
<td>Doesn’t have P</td>
<td>.01 x 19,900 = 199</td>
<td>.99 x 19,900 = 19,701</td>
</tr>
</tbody>
</table>
The standard interpretation

- Probability questions like these have been posed to groups of highly educated people, physicians for example, who would have a fairly high stake in getting them right. In a similar example on p. 122 B&T note that Harvard medical faculty and students routinely ignored the base rate and wildly overestimated the probability that a person who tested positive had the disease.
- The standard interpretation of these sorts of results is that even highly educated people don’t reason well about probability.
- Of course, one can reply to this that even highly educated people are not highly educated in matters of probability, and that we would expect more accurate answers if they were. This is right, but note that it really isn’t a defense of the view being challenged here, namely that human beings are rational by nature. If being rational by nature means having a predisposition to take relevant presented evidence into account, then we would have to be saying that these educated people are completely unaware that the base rate would be relevant.
- An analogous case we’ve discussed in the past is confirmation bias: the tendency to persist in a view based on the relatively few confirming instances of that view as compared to the preponderance of disconfirming instances. Educated people are generally aware of the relevance of disconfirming instances, yet they ignore them quite predictably.
Do people make reasoning errors?

- B&T look at two arguments that base rate neglect and other common mistakes are not actually errors in reasoning because people do not actually make errors in reasoning. This may strike you as a strange position, but it is actually pretty common.

- Consider a formal fallacy like affirming the consequent.

\[
P \rightarrow Q \\
Q \\
P
\]

- This is obviously deductively invalid, and we can safely say that anybody who reasons according to this pattern is making an error in reasoning. However, the question remains whether anybody actually does reason according to this pattern.

- If, for example, I were to say: “Sam must be a dog because Sam has fleas and all dogs have fleas,” this could be reconstructed as:

If something is a dog, it has fleas.
Sam has fleas.
So, Sam is a dog.

- This clearly has the structure of affirming the consequent. But is it reasonable to attribute that reasoning to someone? Isn’t it always more plausible to say that no matter what they said, they really didn’t mean that?

- This is a long discussion, but the basic idea is just that whenever we interpret what people say we always presume their rationality. In other words, the assumption of rationality is a basic rule of interpretation. Of course, one reasonable response to this is that rules have exceptions.
An analogy

According to what Leibniz called the Principle of Sufficient Reason (PSR), everything happens for a reason.

This principle is normally understood as applying to the explanation of nature. The idea is that project itself simply presupposes that at some fundamental level nature makes sense. To say that things happen for no reason is simply to give up on the project of understanding it.

The PSR can also be applied to people and the way they think. Here, however, we have a bit of ambiguity to resolve. According to the standard interpretation of the principle, a reason is basically a cause. So if we apply it to explaining the behavior of people, then we are dealing with them as entities that can be explained in causal terms.

However, we can also apply it to people insofar as they are rational agents. In this case, the PSR would be asserting that people as rational agents always make sense at some fundamental level. To assert otherwise is to simply give up on the project of understanding them.
Performance vs. Competence

- But can’t we believe in the PSR while allowing that people still sometimes make random ass mistakes?
- The answer to this is yes. Noam Chomsky, the renowned linguist, articulated the performance/competence distinction to make sense of this fact. The basic idea here is that we rarely perform at the level of which we are theoretically capable. Our competence is normally manifested in our ability to detect errors when other people make them. We may make the very same mistakes ourselves without realizing it, but when they are pointed out we can recognize them. (Think, for example, about how easy it is to recognize the poor writing of others but how difficult it is to do better.)

- So, of course, those who defend the idea that you can’t give evidence of a person’s irrationality will want to say that poor reasoning performance is not evidence of irrationality. And, of course, that people are fundamentally competent is simply axiomatic.

- The performance/competence distinction is useful, but it can definitely be overdone. From a naturalistic point of view, the claim that a person is competent at any activity should be falsifiable at some level. For example, if somebody has been taught the problem of confirmation bias, and yet continues to reason from instances to generalizations without ever concerning himself with disconfirming instances, then at some point, and at some level, it seems like we have to doubt their competence.

- It may be, for example, that they really do understand the problem, but because they have other emotional or ideological issues, they simply can not prevent themselves from reasoning in this way.

- Why wouldn’t we call that evidence of irrationality?
Gigerenzer’s conceptual argument

- Gigerenzer argues that Bayesian problems like the one we considered earlier can’t provide evidence of human irrationality because people simply do not regard questions about the probability of particular events to be meaningful.
- Recall that the question in the previous example is: What is the probability that you have philosophitis given that you have tested positive?
- Gigerenzer believes that to most people it really does not make sense to talk about the probability of a particular event. According to what is known as the frequentist interpretation, the probability of an event is just a measure of the frequency with which it has occurred in the past. So if we say that there is a 90% chance that you have P, given that you’ve tested positive, that can only mean that 90% of those who test positive actually have P.
- So, the point is that unless we are careful to express problems in a frequentist vocabulary, most people will find the questions very confusing and we should expect a high error rate.
- B&T basically think that Gigerenzer is wrong about this. They argue (p.127) that if you receive a positive test for P and you understand that the test has a margin of error, you will naturally ask: How likely is it that I actually have P?
- In fact, it seems that ordinary people are capable of thoughts like this: “I don’t care how frequently other people who got a positive test turned out to have P. I’m asking about me!”
Cohen’s conceptual argument

- Cohen argues that it is incoherent to find normal human adults to be incompetent at reasoning. He acknowledge that they sometimes perform poorly, but that their competence is simply an assumption of interpretation as we explained earlier.

- With specific regard to reasoning about probability Cohen argues that we have to begin by accepting as accurate the intuitive judgments of probability made by laypeople. Of course, finding out what these are is a delicate matter, since at any given time an judgment might be a performance error.

- This view leads Cohen to argue that when base rate neglect is actually occurring- and he agrees that it does- it is rational. This sounds pretty crazy and B&T are duly scandalized by it, but here is one way to try to justify it.

- B&T consider (p.132) a situation in which someone in a very low risk group tests positive for HIV. In fact, the likelihood that such a person has HIV after one such test is very low, so we might think it is rational to conclude this. But someone might argue that it is actually rational to attach a much higher probability based on a precautionary, better-safe-than-sorry approach.

- More practically rational, that is, to operate with an exaggerated idea of the danger than a correct idea that might encourage risk taking behavior.
The precautionary principle

- The precautionary principle (PP) is the claim that when we are uncertain about the dangers associated with a proposed course of action A, and there are plausible reasons for thinking the risk of doing A may be great, we shouldn't do it. This sounds very sensible, but as stated it completely fails to take into account the dangers associated with not doing A.

- So, for example, a country confronted with the very real danger of mass long-term famine might be well-advised to consider the use of genetically engineered drought-resistant crops. In this case, a bias toward exaggerating the danger of these technologies may be ill-advised.

- More relevant to the criticism of Cohen’s argument, an exaggerated sense of the danger of something may be considered irrational because it can permanently impair our ability to grasp or even desire to know the actual danger.

- To return to the AIDS example, people who have been diagnosed with HIV aren’t immediately prone to do constructive things on that basis of that diagnosis. Some have committed suicide, others have stopped worrying whether their sex partners were infected.
Irresolvable differences?

- It is always a temptation to conclude that these sorts of debates are either entirely semantic or reflect an irresolvable degree of subjectivity. But this is something we are prone to conclude too quickly.
- Often, for example, we find ourselves thinking that radically different cultural practices are evidence of very different moral views, only to find out that when empirical conditions change (e.g., significant increases in prosperity and freedom of movement) the moral intuitions begin to converge.
- In the case of rationality, then, it is really important to ask whether what appear to be fundamentally and irresolvable different intuitions about what is rational in any given case really are.
Intuitions about base rate

- For example, Gigerenzer argued that base rate neglect may be seen as rational from the frequentist perspective, and the assumption that there are irresolvable differences between frequentists and Bayesians would explain differential answers to low base rate problems.
- But this explanation can be tested. It would predict, for example, that the typical person who ignored base rate and reported that base rate was actually irrelevant, would tend to gravitate to a well-constructed argument for ignoring base rate rather than a similarly constructed argument for paying attention to it.
- B&T consider studies done by Stanovich (p. 134) in which the subjects who believed base rates to be irrelevant were isolated and queried.
- The example they were asked to consider was a fictitious disease (Digirosa), so we might as well stick with Philosophitis. So suppose we remind subjects that Philosophitis is an extremely rare disease, and ask them whether that is relevant to determining whether to believe the results of a test.
- We then isolate the group that said no, and then ask them to consider two pieces of reasoning.
Intuitions about base rate 2

- The examples are as follows:
  - The percentage of people with Philosophitis is needed to determine the probability because if Philosophitis is very rare in the population and some people without Philosophitis also test positive, then the probability of having Philosophitis may be low even if you’ve tested positive.
  - The percentage of people with Philosophitis is irrelevant because this particular person did test positive and thus the percentage of people who have Philosophitis is not needed when trying to determine the probability that someone has Philosophitis if they’ve tested positive.

- The question is how do these arguments affect people when they are exposed to one or both of them.
- Stanovich reported that when both arguments were given to the subjects, those who changed their mind were more likely to change their minds in the correct direction.
- Stanovich also reported that the correct argument was more effective in changing the minds of people who ignored base rate, then the bad argument was effective in changing the mind of people who paid attention to it.
- While evidence like this tends to contradict Gigerenzer, in some ways it may support Cohen. How, after all, do you explain the tendency to change ones mind in the right direction without presupposing people’s fundamental rationality?