Probability, Specificity, Accuracy, Fairness

Marcello/Rafal

1 The Problem

Does aiming to offer more specific stories (theories, accounts, narratives, explanations), as opposed to merely highly probable stories, lead to more accurate and more fair decisions? This is a generic question and needs to be made precise. It is inspired by a remark made by Popper about science:

> Science does not aim, primarily, at high probabilities. It aims at a high informative content, well backed by experience. But a hypothesis may be very probable simply because it tells us nothing, or very little. A high degree of probability is therefore not an indication of goodness. (Popper, 2002, p. 416, Appendix *IX)

Specificity here is defined as informative content. There is an inverse correlation between informativeness and probability. Other things being equal, the more probable a statement, the less informative the statement. A disjuction $A \vee B$ is less informative than a single disjunct, but the disjunction is more probable than the single disjunct.

So how can a policy of adopting more informative (more specific) theories (stories, narratives, etc.) lead to better accuracy? It would seem to be the opposite—that someone who adopts less informative theories is more likely to be right, other things being equal, since informativeness and probability are inversely proportional.

But this depends on how we define accuracy. Popper's intuition must be understood against the background fact that theories are challenged, scrutinized, falsified. So, the claim is that, between a specific theory that has been challenged and survived challenges and a non-specific theory that has been challenged and survived challenges, the specific theory is more likely to be right, more credible, or something to that effect.

This may be right. But how do we show that it is right?

2 Simulating specificity and accuracy

To test Popper's intuition we can use a computer simulation.

2.1 Basic set up

- Suppose we have a plane consisting of 1000 points (or more). A maximally informative description of how the world is like is an assignment of 0's and 1's to each of these points. This is the maximally informative or maximally specific theory.
- Less informative theories will leave undecided assignments of 0's and 1's to certain points. The more undecided points, the less informative the theory. The theory that leaves all points undecided is the least informative, corresponding to a tautology.
- The actual world is a maximally specific assignment of 0's and 1's to all the points. The goal in science and trial proceedings is to to find out what the actual world is like.

- One side, the prosecutor, puts forward a theory—a certain assignment of 0's and 1's to the points. The theory need not be maximally informative. In fact, it never is, but it does have some degree of informativeness.
- The other side, the defense, will proceed to attack this theory. How does one do that? We can imagine that the defense has available resources to seek evidence. Evidence simply consists in information that unravels which points are 0's and which points are 1's. Evidence is partial and fragmentary and reveals only some points. Let's assume for now the evidence isn't probabilistic. It unequivocally and truthfully tells us which points are 0's and which are 1's with no error whatsoever. Let's also leave out misleading evidence. All these complications will have to be addressed, but one thing at a time! The difficulty is to find out such good, truth-conducive evidence.
- Evidence that conflicts with a proposed theory falsifies the theory, and evidence that does not, corroborates it. Comparing bits of evidence with a proposed theory is the adversarial testing that goes on at trial. This process is limited by available resources and time. It cannot go on forever. It will end at some point.
- We can think of the following rule of decision: if a theory is falsified, reject it; and if the theory is not falsified (within a period of time), accept the theory.

2.2 Questios reformulated

- So the question can now be stated more precisely: Suppose Policy A imposes that only very informative theories be discussed at trial and contrast this with Policy B that allows less informative theories. Do trial systems that adopt Policy A end up accepting theories that are false more or less often than systems that adopt Policy B? And what about rejection of true theories?
- Suppose a more specific theory has resisted all challenges and so did a less specific theory. Do we have good reason to believe one more firmly than the other? It would seem that we have better reason to believe the more specific theory, but why exactly? Would accepting the more specific theory lead to accepting a true theory more often than accepting the less specific theory?
- Several scenarios to compare. The proposed theory is true, versus the proposed theory is false. Is a more specific theory (that is false) more quickly be proven false (=rejected) then a less specific theory (that is false)? Perhaps so. What about a more more specific theory (that is true) compared to a less specific theory (that is also true)?
- The answer to these questions may depend on time constraints. If there were not time constraints, perhaps the two policies would be indistinguishable. Is this right? Maybe there is an interesting relationships between time, specificity and accuracy.

2.3 Possisble answers

- Give limited time, it would seem that, under this simple model, the more specific the theory being tested, the more likely it will be proven wrong (=rejected) if the theory is actually false. For imagine that the theory is maximally informative and it is wrong. Revealing one single piece of evidence that conflicts with the theory could be enough to prove the theory wrong.
- Anyway these intuitions should be substantiated by the simulation. I am already getting confused as to what exactly is going to happen.

2.4 Complications

- So we can ask many questions about the credibility of more or less specific theories in the simple set up already. The next step would be to introduce the possibility of misleading evidence, or the allocation of uneven resources between prosecutor and defense. Do these complexities change the general picture?
- The goal here would be to get to a progressively more complex and realistic picture of legal decision-making.

3 Specificity and fairness

- Is there a relationship between fair decisions and specificity? Perhaps one way to tackle this is to compare trials in which innocent defendants have available less resources or less time. This presumably has a negative impact on accuracy to their detriment. This can be checked with the simulation. Seems obvious but good to check!
- In terms of specificity and fairness, one thing to check could be whether more specificity can reduce the gap in accuracy between trials in which defendants have more resources and more time compared to trials in which defendants have less resources and time? In other words, is specificity a decently good corrective measure to compare for gaps in accuracy between defendants with more and less resources and time?

4 Probability, calibration, specificity and accuracy

- One question that we leave out is whether using probability theory, Bayesian networks, etc, has benefits for accuracy. How can this claim being tested? It is not obvious.
- The connection might be that probability theory helps to arrive at well-calibrated judgments about the probability (credibility) of theories. One strategy could be to see what happens if one's probability assignments are mis-calibrated. What happens if the fact-finders are over-confident or under-confident in their judgments about the credibility of a proposed theory? Does mis-calibration hamper accuracy? And if so–it seem obvious that it does—how exactly?
- A connection with the specificity bit is that, perhaps, aiming to test more specific theories is more likely to lead to calibrated judgments about the credibility (probability) of these theories compared to less specific theories. Is this right? How do we model that?