Why favour simplicity?

ROGER WHITE

Among theories which fit all of our data, we prefer the simpler over the more complex. Why? Surely not merely for practical convenience or aesthetic pleasure. But how could we be justified in this preference without knowing in advance that the world is more likely to be simple than complex? And isn't this a rather extravagant a priori assumption to make? I want to suggest some steps we can take toward reducing this embarrassment, by showing that the assumption which supports favouring simplicity is far more modest than it first seems.

1. Favouring simpler behaviour

We extrapolate from observed regularities with some measure of ignorance concerning the causal mechanisms which give rise to them. It may be helpful, therefore, to consider how we ought to theorize about a certain idealized model of a regularity producing mechanism, and then see how well this serves as a model for ordinary physical systems.

Consider a black box with a rotating dial and pointer. Movement of the dial results in movement of the pointer by some unknown mechanism inside the box. We can think of the box as computing a function from dial positions to pointer positions. Turning the dial and reading the pointer we collect data points, and from these we try to make an educated guess as to which function the machine computes. Suppose that the dial

ANALYSIS 65.3, July 2005, pp. 205–210. © Roger White

and pointer each have a large but finite number of positions they can be in, so there is an enormous but finite number of possible functions to consider. What I want to consider here are the assumptions that are required to support the following Simplicity Favouring Principle:

SFP: The greater the complexity of a function *f*, the less probable it is that the machine computes *f*.

There are two kinds of complexity to be distinguished here: the complexity of the inner workings of the machine, and the complexity of its outer behaviour: the complexity of the function that it computes. The complexity of the mechanism in the box is roughly a matter of how many different kinds of parts it contains, all of which are intricately linked so that a change in any part would make a major difference to the workings of the mechanism. The notion of complexity of a function is harder to characterize, but I will assume that we recognize it well enough when we see it. For instance a linear function is intuitively characterized as simple, and higher order polynomials as more complex.

The intuitive complexity of a function bears an important relation to the complexity of a mechanism: the more complex the function, the more complex the mechanism required to compute it. A simple mechanism consisting of just a rubber band connecting the dial and pointer could only compute a simple linear function; a more complex function like $7x^3 - 6x^2 + 2x + 1$ can be computed only by a machine with very many different moving parts connected in an intricate way. An appropriate way to measure the complexity of a function then, is in terms of the simplest machine required to compute it. Let's formalize this by letting the complexity function *c* map each possible function onto the least degree of machine-complexity required to compute it. Now letting

F = the machine computes the function *f*

the thesis under consideration can restated as

SFP: P(F) decreases as *c*(*f*) increases.

Let's suppose that we have a rough numerical measure of machine-complexity:

M_i = the mechanism in the black box is of complexity *i*

Now the probability that the machine computes the function *f* can be given as

$$(1) \quad P(F) = P(F|M_1)P(M_1) + P(F|M_2)P(M_2) + \ldots + P(F|M_n)P(M_n)$$

where *n* is the maximum degree of complexity for a machine that could fit in the box. A machine cannot compute a function whose complexity measure is greater than the complexity of the machine, i.e. $P(F|M_i) = 0$ if $c(f) > i$. So

(1′) $P(F) = \sum_{i=c(f)}^{n} P(F|M_i)P(M_i)$

The following fact is now crucial: The number of functions that a machine can compute increases rapidly with the complexity of the machine. Given only few bits and pieces, there is a very limited number of ways that I can arrange them to causally link the dial and the pointer. Given dozens of gears and pulleys and levers and whatnot, the possibilities are much, much wider. As result,

(2) $P(F|M_i)$ decreases rapidly as i increases.

A consequence is that the only way to *avoid* the result SFP is to assign a prior probability distribution over the hypotheses $M_1 - M_n$, such that $P(M_i)$ *increases* dramatically as i increases. That is, the only assumption we need to support SFP is the relatively modest one that

(A) The machine is not considerably more likely to be complex than simple.

It is not obvious exactly which possible probability distributions over $M_1 - M_n$ are rationally acceptable, so it might seem wise to remain entirely noncommittal. But this is not always an option. We may just be curious as to what the box will do next, or may be forced to act in ways whose outcome will depend on its behaviour. (We often need to make the best estimate we can as to how the world will behave.) So we are forced to distribute our credence in some way among the possible degrees of machine-complexity. From a position of ignorance about the contents of the box, a distribution that violates (A) would seem rather arbitrary and unreasonable. What could justify someone in having a *much greater* expectation that the box is chock full of intricately linked moving parts, than that it just contains one or two? In any event, (A) is a more modest assumption than that the mechanism is especially likely to be simple. Someone sceptical about our ability to judge a priori the degree of complexity of a black box's contents might assign a roughly even initial credence over $M_1 - M_n$. By the argument above, this would be more than enough to support SFP.

Someone may object to my appeal to (A) as follows. There are more ways for a mechanism to be complex than simple (indeed the argument above depends on this fact). So a random selection of a specific machine-design is most likely to yield a complex mechanism. From a position of ignorance we should treat the box's mechanism as if it had been randomly selected from the possible machine-designs. Hence contrary to (A), the box is more likely to contain a complex mechanism.

In response to this complaint, we should keep in mind that the main ambition of my argument was to the reveal that the assumption (A) that is required to support SFP is more modest than it first seems. Rather than having to suppose that the box's mechanism is especially likely to be simple, we need only assume that it is not considerably more likely to be complex. This can make our preference for simpler functions seem less like an arbitrary and unwarranted bias. For on the face of it, (A) appears rather innocent. Here I am simply appealing to our shared intuition. What is important to note here is that our intuitive judgment of the plausibility of (A) is in no way *derived* from our preference for simpler functions. Quite independently of any thoughts about what we should expect the box to *do*, it seems odd to expect the box to contain a very complex mechanism rather than a simpler one. To see that an assumption (A) that we find plausible independently of SFP, turns out to entail SFP, is to make some philosophical progress when the credentials of SFP are in question.

This progress will not be very satisfying though, if the stated objection gives us a good reason to abandon our natural acceptance of (A). I don't see how it can. Of course if we happened to know that the machine was constructed by first picking a specific machine-design out of a hat and implementing it, then we should divide our credence equally among the possible machine-designs and the functions they generate, thus violating SFP. But of course we typically know no such thing. And without such a background assumption, this distribution of credence seems inappropriate. Consider an analogy. For some odd reason, I have to try to set reasonable betting odds on the number of children that the Joneses have. Maybe they have just one, but perhaps as many as a six. Ignoring what little statistical data we have to go on, it is hard to know what to expect. But it doesn't seem reasonable to strongly bias one's estimate toward very many children rather than just a few. Someone might, however, argue as follows. As there are more specific ways to have lots of kids, they are more likely to have many. That is, instead of just partitioning the possibilities in terms of the number of kids, we could divide them further in terms of specific kid-arrangements. The more kids there are, the more kid-arrangements there are. There is a very limited range of ways to arrange a single child. But if there are six of them, there might be two in their bedrooms, three in the yard, and one in the kitchen, or two on the front porch, and the rest in the tree-house, or all six in the attic, or ... Since the number of possible kid-arrangements rises dramatically with the number of kids, if we assign a roughly equal credence over the possible kid-arrangements, we should be willing to bet at considerably higher odds on their having very many kids, than just a few.

Now it appears just obvious that this is the wrong way to go about predicting how many children the Joneses have. Whatever our initial estimate, we shouldn't drastically raise it in response to the argument above. We ought to be similarly sceptical about the suggestion that we should distribute our credence evenly over specific machine-designs as a way of challenging (A).

2. Favouring simpler mechanisms

We also favour simplicity among mechanisms. Whatever our initial hunch about the box's contents, once we have collected a body of data we are more inclined to accept the hypothesis which attributes it to the simplest mechanism. Call this the Simple Mechanism Favouring Principle (SMFP). Let D be a list of the several data points observed, and f be the simplest function fitting these points.

SMFP: If $c(f) \leq i$, then $P(M_i|D)$ decreases as i increases

(Only if $c(f) \leq i$ is $P(D|M_i) > 0$, i.e. if the mechanism is sufficiently complex to compute a function fitting the points of D.) Now,

(3) $$P(M_i|D) = P(D|M_i)\, P(M_i)/P(D)$$

So $P(M_i|D)$ is proportional to the product $P(D|M_i)P(M_i)$. As we've noted, the range of functions computable by a mechanism increases rapidly with the mechanism's complexity. So, likewise, the number of possible observed data-point sets increases rapidly, and hence

(4) If $c(f) \leq i$, then $P(D|M_i)$ decreases rapidly as i increases.

Once again, one must violate assumption (A) that $P(M_i)$ does not dramatically increase as i increases, in order to avoid the result SMFP. So (A) supports the policy of attributing observed regularities to simple mechanisms.

3. Application to real systems

So much for our idealized model. How does this apply to problems in the real world? One simplification we made was to suppose that the dial and pointer could take finitely many positions, when in principle they may move continuously, creating uncountably many possible behaviours. This simplification seems harmless. Ideally as we more finely discriminate the dial and pointer positions, increasing the number of possible movements without bound, the very same result will obtain. So the probability density function over the continuum of possible machine-functions should be biased toward the more simple. Of course it is doubtful that such precision even makes sense for a real nuts-and-bolts machine (or any real physical system). At some point there becomes no fact of the matter as to whether a system's behaviour can correctly be described by, say, f(x), rather than by $f(x) + 10^{-100000\ldots}$.

So while the situation is not nearly so clear cut, the two factors which were crucial to the argument above appear to hold in ordinary cases: (i) the more complex the hypothesized regularity, the more complex the causal mechanism required to produce it, and (ii) the range of regularities which could be produced by a causal mechanism increases dramatically with the complexity of the mechanism. So, as in the idealized case, relatively modest assumptions about the likely complexity of the causal mechanisms at work in the world support the favouring of simple hypotheses over complex ones. Perhaps this does not nearly solve all the puzzles concerning the role of simplicity in theory choice. But it provides a significant insight which can ease our discomfort about favouring simplicity.[1]

New York University
New York, NY 10003-6688, USA
roger.white@nyu.edu

[1] Thanks to Adam Elga and Jenann Ismael for discussion on this topic. I've floated these ideas in classes at NYU and in comments on Jenann's 'Why simpler theories are more apt to be true' at Metaphysical Mayhem VI, Syracuse 2001. I'm grateful for the feedback I've received on those occasions.

ANALYSIS 65.3, July 2005, pp. 210–18. © Cian Dorr