Knowledge, Trade-Offs, and *Tracking Truth*

Peter Godfrey-Smith

Harvard University

1. Introduction

There are so many ideas in Roush's dashing yet meticulous book that it is hard to confine oneself to a manageable menu of issues. Here I will reluctantly omit discussion of scientific realism, fallibilism, contextualism, and much else. My focus will be on one core part of Roush's account of the "tracking" of facts, and the relation between tracking and knowledge.

In approaching rival "theories of knowledge" it is helpful to distinguish three sets of facts. An imaginary complete philosophical theory in this area would have things to say on three different matters. First, there are facts about our ordinary concept of knowledge, and related concepts such as those of evidence and justification. The term "concept" might (should) be suspicious here, and I use it in a low-key sense. Perhaps it is better to imagine a complete account of how the words "knowledge" and "know" (etc.) are used, in both everyday circumstances and special ones. A second set of facts concerns how we human beings, and perhaps other organisms, are actually connected to the world with respect to epistemic matters. What sort of coordination and contact with external affairs do our perceptual states, beliefs, and theories achieve, when things go well and when things go badly? What sort of outcomes can be reasonably *hope* for, when we set out on the task of representing the world? Philosophers have to use *some* terms and concepts when developing this second body of theory, but they need not be familiar, everyday terms like "knowledge." A third set of facts follows from, and concerns the relation between, the first two. How do the terms, concepts, and pictures that people ordinarily employ in this area relate to our actual epistemic attributes?

This three-part breakdown is not peculiar to knowledge. The same could be said for philosophical work on causation, or belief. There are facts about ordinary attributions

of causation, facts about how the world runs, and facts about how our concept of cause relates to the running of the world. In the case of both causation and knowledge, philosophical work tends to move between these three sets of facts, sometimes with acknowledgment and sometimes without.

Roush makes "tracking" central to her account of knowledge. Given the framework above, there are several ways that tracking might be important in this area. Tracking might be part of an account of the ordinary concept of knowledge, whether or not the world contains cases of tracking in Roush's sense. Alternatively, the idea of tracking might be a useful tool when describing actual connections between beliefs and world, whether or not ordinary people realize this. As I read Roush, she thinks that tracking has both these roles; the ordinary concept of knowledge has succeeded in picking out a natural epistemic kind, and tracking is that kind's nature. So there is a harmonious relation between our everyday concepts and how things actually work, with respect to this issue. As a consequence, Roush moves readily between discussions of intuitions about what *counts* as knowledge, by normal standards, and discussions of the role of tracking in the natural world itself – how it might figure in an epistemology that goes far beyond folk intuitions. I will discuss each of these in turn.

2. Tracking and the Ordinary Concept of Knowledge
Roush's notion of tracking is descended from Nozick's (1981). The heart of Nozick's account comprises two criteria expressed with subjunctive conditionals. As re-formulated by Roush (p. 42) they are:

(3) If p were not true then S would not believe that p.
(4) If p were true then S would believe that p (and would not believe that not-p).

As Roush says, there is a special understanding of the "if... then..." assumed here, so that (4) does not follow from the mere fact of truth and belief. (I omitted the asterisks Roush uses to indicate this). Roush's theory takes this notion of tracking as raw material, but makes two modifications. One is giving the analysis of knowledge a recursive structure,

so that knowing that p requires *either* tracking the truth of p *or* correctly inferring p from other facts that one tracks. The other modification is using conditional probabilities instead of conditionals. So Nozick's (3) and (4) are replaced by III and IV below.

III. $P(\sim b(p)|\sim p) > t$, and

IV. $P(b(p)|p) > t$, and $P(b(\sim p)|p) < 1 - t$.

Here "b(p)" is read as saying that p is believed (by S), and t is some high threshold that Roush usually treats as 0.95. In these comments I will not worry about the interpretation of the probabilities, the fallibilist setting of t, or the need for the second clause within IV. My focus will be on the role of things *like* Roush's IV.

Requirements akin to (3) are found in many other views of knowledge, including reliabilism (Goldman 1986), Dretske's informational view (1981), Armstrong's nomic dependence view (1973), contextualist relevant-alternative views (Lewis 1996), Sosa's "safety" view (1999), and others. I emphasize the "akin"; sometimes the connection asserted is one between an agent's evidence and a fact, not between belief and fact, and there are many other variations. But the common theme is clear: the agent would not have believed that p, or received the evidence on which they based their belief that p, if p were not true. There is a clear difference between this first family of ideas and the one expressed by Nozick's (4) and Roush's IV. This is the idea that a belief that p is not knowledge if it is not the case that *given* that p is true, the agent will believe it. In Roush's formulation, the probability that p is believed, given that it is true, is high. (Again, I am not going to worry about the second requirement in IV, just the first.) There would be other ways of expressing requirements like IV, but the heart of it is the requirement that the agent is robustly, probably, or reliably *going to notice* the truth of p, in a range of circumstances other than the actual one where, by assumption, the agent does believe that p. So what is ruled out from being knowledge is, for example, a case where the agent is so inattentive that for most of the ways that p could be true in those sorts of circumstances, the agent would not register it. The agent *did* register p in the actual case, but they could easily have missed it. If that is so, then their actual belief – the one they were lucky enough to get to – cannot count as knowledge.

Many writers within the externalist/reliabilist/relevant-alternatives tradition have not required anything like IV. Dretske, as far as I know, does not raise the possibility at all. Goldman (1986) discusses a related measure ("power") but keeps it separate from the ingredients for knowledge. Armstrong raises the possibility of a IV-like clause and says that it should not be included.

> Given that a reliable thermometer is registering "T^O," then the environmental temperature is T^O. But, equally, given that the environmental temperature is T^O, then the good thermometer will register "T^O." Now this second feature is not necessary for non-inferential knowledge. (1973, p. 174)

I think the no-IV camp is right on this point; this is not part of the ordinary concept of knowledge. Talk of knowledge involves a particular mode of epistemic assessment. We assume the agent has some belief, and we assess the belief's credentials. We don't have to ask how likely that psychological state was to arise, given its truth. We *can* ask that; we can ask whether the agent was "lucky to notice" some fact, as opposed to there being "no way they were going to miss it." But if the person was lucky to register the fact, that does not mean they can't know it, given that they did register it.

Though I side with the no-IV camp, Roush's view has fewer problems here than Nozick's account. This is because Roush's theory allows knowledge of one fact by inference from other facts that are tracked. The advantage this gives Roush in this context is seen in her discussion of the Jesse James case (pp. 68-71). Roush says there can be cases where a person is lucky to come across facts that are decisive evidence for something, but the person's relationship to these evidentially useful facts is one of tracking, and they can know more by making inferences from what they saw. Betsy is lucky to see Jesse James' mask slip as he runs out of the robbed bank. Because she is lucky, her belief that the robber was James does not directly track the fact that it was James. But her beliefs about seeing a man with a certain appearance running out of the bank do track those facts (I will grant this to Roush), and from those beliefs she infers that the robber was "very probably" James. This reply only works, however, when the two-step story involving inference from other facts is independently plausible. And a failure of IV-type tracking relationships can occur in the case of non-inferential

4

knowledge. We can see this even with Betsy. Assume she is usually so absorbed by her iPod that she notices hardly anything around her, and just by chance happened to see the running man with the slipping mask. Then she does not track (in Roush's sense) the facts used to infer that the robber was James.

At this point we should look at Roush's explicit argument for the need to include something like IV (p. 39) Roush puts it like this.

> Suppose that the president of the United States believes that Bob Woodward is a good journalist, and suppose that Woodward is a good journalist.... [But] the president believes that Woodward is a good journalist because of nothing except the fact that Woodward published a flattering book about him. ... The tracking view has a good explanation of why we take the president to lack knowledge. The second tracking condition [(4)/IV] says that the belief of a subject who knows p must be such that if p were true but other things slightly different, then the subject would believe p.... One way to fulfill the antecedent of this conditional for our case is through a scenario in which Woodward is a good journalist and he publishes an *unflattering* book about the president in question.

Then, Roush supposes, the president would not believe that Woodward is good, and this shows that in the actual case the president failed IV. I see the case differently. I accept that we are reluctant to say the president knows that Woodward is good, but that is because of a different possibility, the possibility in which Woodward is a *bad* journalist but publishes a flattering book about the president. The way the president is described here, this is a case where the president would still believe that Woodward is good. So the president is failing on a III-type test, as well as a IV-type one. The president's belief is not one that would not exist were it to be false. Given that this case includes a III-failure as well as a IV-failure, the absence of knowledge in the Woodward/president case does not show the need for a IV-type clause.

So I think that IV is not a real part of the everyday concept of knowledge. However, that does not mean that IV has no importance at all. At this point Roush might reply that IV is a natural complement to III, as far as epistemic goodness goes. Epistemic success involves both features; the belief is non-accidentally true, and also non-accidentally *believed*, given that it is true. A failure with respect to IV is an epistemic failure. One way to see this is via the link between tracking and "power," in the Baconian

sense of technological control, which Roush rightly emphasizes. True belief is valuable, but tracking the truth is more valuable, and tracking in the III+IV sense is more valuable than tracking with respect to III alone.

I think that all those claims are right. The argument of this section is that those facts are not reflected in the ordinary concept of knowledge.

3. Tracking Outside of Knowledge

Putting aside questions about ordinary concepts, let us turn to another side of Roush's project: giving a theory of our actual epistemic attributes which connects to the explanation of Baconian power. I agree with Roush that the idea of tracking facts is a useful one with which to build such theories and models, and agree that something like the criteria expressed in her III and IV are appropriate in giving the idea of tracking more precise content. In this section I will argue that Roush may give insufficient weight to a particular feature of the relations between III-like and IV-like properties. This is the fact that, in many circumstances, there is a trade-off between these two desirable relations that one's inner states can have to the world.

The idea of such a trade-off has been recognized in a scattered tradition within epistemology. William James expressed one version of the idea in "The Will to Believe."

> There are two ways of looking at our duty in the matter of opinion – ways entirely different, and yet ways about whose difference the theory of knowledge seems hither to have shown very little concern. *We must know the truth*; and *we must avoid error* – these are our first and great commandments as would-be knowers; but they are not ways of stating an identical commandment.... [T]hese, we see, are two materially different laws; and by choosing between them we may end by coloring our whole intellectual life. (1897, pp. 99-100)

For James, an agent faced with decisions about epistemic policy, must make a trade-off between two incompatible goals. One can studiously avoid false beliefs, by using cautious methods of inference which yield a high ratio of truths to falsehoods. Alternatively, one can take a more adventurous approach, making commitments on the basis of slimmer evidence, hoping to believe more truths even if this entails the risk of accepting some falsehoods as well (see also Levi 1967).

Within James' discussion, *avoiding error* corresponds to maximizing $P(\sim b(p)|\sim p)$; it is ensuring that one does not believe things when they are false. And *knowing the truth* is maximizing $P(b(p)|p)$; it is ensuring that one believes things when they are true. (Roush might see James' choice of words here as putting pressure on some claims in section 2 above.) In an earlier discussion (1996) I used the term "Jamesian reliability" for (roughly speaking) $P(b(p)|p)$, and "Cartesian reliability," not for $P(\sim b(p)|\sim p)$ but for the related "safety" property, $P(p|b(p))$.[i] Both Descartes and James did care about both kinds of epistemic success, but each famously oriented their epistemology around one particular goal: avoiding error, in Descartes' case, and registering hidden but important truths, in James'.

The idea of a trade-off of this kind is dependent on a number of assumptions, some probably shared by Roush and some not. If an agent is forced to either believe that p or believe its negation, there is no real trade-off. Then failing to believe a truth involves believing a falsehood. The trade-off requires the possibility of some sort of third option, either suspension of opinion, obliviousness, or some other way of "doing without" a belief (James' term) without believing its negation. Also, all this talk of the probability of something "being believed" assumes a framework in which a qualitative notion of belief makes sense, as opposed to all opinion being a matter of degree. Roush is cautious on this issue, accepting an underlying quantitative concept of belief but treating a degree of belief over 0.95 as close enough to simple belief (p. 48).

The trade-off I am discussing here is reflected in the distinction between Type I and Type II errors in classical statistics. The place where it has been modeled in the most detail, however, is probably "signal detection theory," which is used in psychophysics and behavioral ecology. (Green and Swets 1966). In this model, an agent's task is to detect a state of interest in the world, given the presence of "noise" which makes available sensory cues ambiguous. In the simplest case, some particular observable variable, X, is used to decide whether the signal is present or not. Lower values of X tend to be associated with noise (S_1), and higher values with signal (S_2), but the distributions giving the probability of observing various values of X given signal, and the probability of observing various values of X given noise, overlap as in Figure 1. There are two ways for the agent to be right and two ways to be wrong. A "hit" is a judgment that signal is

present when it is. A "correct rejection" is a correct judgment of mere noise. A "false alarm" is a judgment that signal is present when there was only noise, and a "miss" is a failure to detect the signal when it was present. So if the classical statistics terminology is used and noise is the "null hypothesis," then a false alarm is a Type I error and a miss is a Type II error.
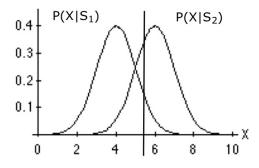

Figure 1: Epistemic life from the point of view of signal detection theory

Signal detection theory supposes that these various good and bad decisions are associated with payoffs, and that there are also overall or "prior" probabilities for signal and noise. Then it is possible to work out, for a given agent, the optimal setting of the threshold value of X (illustrated with a vertical line on Figure 1), so that if the agent sees a value of X above that value they should treat the situation as one of signal rather than noise (Godfrey-Smith 1996). Mapped onto Roush's framework, and assuming that p corresponds to signal, an agent who sets the threshold value of X far to the right on Figure 1 is pursing tracking with respect to clause III, and sacrificing their score on IV. An agent setting the threshold to the left is pursuing a good score on IV and sacrificing III. An agent can pursue both goals at once only by reducing the overlap between the two distributions. For example, if X is the observed value of some property of a sample, then increasing the sample size may reduce the variance of the two likelihood functions $P(X|S_1)$ and $P(X|S_2)$, thereby reducing the overlap between them and making epistemic life easier. (Signal detection theory measures the overall ease of epistemic life in this situation with d', which is the difference between the means of the two distributions divided by their common standard deviation.)

8

This is a very artificial model of the human epistemic predicament. The cognitive and the behavioral are tied extremely closely together, and the standard description imposes (via the "signal" versus "noise" terminology) an asymmetry in the status of S_1 and S_2 which is also questionable, once we import the model into philosophy. But if we take this model as representing, in highly simplified form, something basic about our attempts to "track" the world, then it does cast the relationship between III and IV into a different form from that seen in Roush. Our scores with respect to III and IV reflect a trade-off that we make within the constraints imposed by the general informational profile of our links to the world. If we want to improve our standing with respect to III, we can *either* make an adjustment within a standing set of constraints (shifting the threshold in Figure 1), *or* we can go looking for a better cue. Roush might say that science is (among other things) a highly organized attempt to pursue this second option.

References

Armstrong, D. M. (1973). *Belief, Truth, and Knowledge.* Cambridge: Cambridge University Press.

Dretske, F. (1981). *Knowledge and the Flow of Information.* Cambridge, MA: MIT Press.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature.* Cambridge: Cambridge University Press.

Goldman, A. (1986). *Epistemology and Cognition.* Cambrige, MA: Harvard University Press.

Green, D. and J. Swets (1966). *Signal Detection and Psychophysics*. New York: Wiley.

James, W. (1897). "The Will to Believe." In *The Will to Believe, and Other Essays in Popular Philosophy.* New York: Longmans.

Levi, I. (1967). *Gambling With Truth.* New York: Knopf.

Lewis, D. (1996). "Elusive Knowledge." *Australasian Journal of Philosophy* 74: 549-67.

Nozick, R. (1981). *Philosophical Explanations.* Cambridge, MA: Harvard University Press

Roush, S. (2005). *Tracking Truth: Knowledge, Evidence, and Science*. Oxford: Oxford University Press.

Sosa, E. (1999). "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13: 141-153.

---

[i] Roush discusses the relation between P(~b(p)|~p) and P(p|p(b)) in Chapter 4, noting that they do not trade off and also coincide in an extreme case; if either is 1 then the other is 1.