Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
Vol. LXXIX No. 1, July 2009

Précis of *Tracking Truth*

SHERRILYN ROUSH
University of California, Berkeley

In *Tracking Truth* I undertook a broader project than is typical today toward questions about knowledge, evidence, and scientific realism. The range of knowledge phenomena is much wider than the kind of homely examples—such as "She has a bee in her bonnet"—that are often the fare in discussions of knowledge. Scientists have knowledge gained in sophisticated and deliberate ways, and non-human animals have reflexive and rudimentary epistemic achievements that we can easily slip into calling "knowledge." What is it about knowledge that makes it natural for us to use the same word in cases that are so vastly different? How is it possible for knowledge to have evolved? What is it about knowledge that it should enhance our power over nature, as Francis Bacon observed? What is it about evidence and knowledge that makes you more likely to have the latter when you have the former? Specialization is necessary to progress, but the division of labor it requires has allowed such questions to fall through the gaps between discussions.

These gaps are opportunities. Sometimes newly discovered problems can bring new and better answers even to old questions. The questions I have asked above are "Why?" questions expressed as (apparently) Socratic "What is?" questions, and that is the approach taken in the first five chapters of this book, to offer explanations of familiar phenomena on the basis of rigorous definitions of knowledge and evidence. One might object that this is an old, not a new, style of answer, and one that I ought to be educated enough to reject. Many have thought the project of giving necessary and sufficient conditions for knowledge was in its death rattle long ago. The most common argument for this conclusion is an empirical one, that no such attempt has ever been successful in giving the right answer for all examples. And when one asks, as one must, what the "right" answer would be answering to anyway, the project can look even more depressing. But even if there is a clear

standard we are consistently imperfect against, and always will be, I think that to view such imperfection as nothing but failure is to fall into the trap of making the perfect the enemy of the good.

It is a mindset that sadly squanders our ideas. Even if we will never succeed in the way specified, it doesn't follow that there is nothing to be gained in the project of formulating general definitions. No scientific theory we know of has ever done a perfect, or even good, job with every case. And this is true from a theory's inception; as Kuhn taught us, every theory is born falsified. I take it we would all hesitate, though, before concluding that this makes scientific theorizing pointless. Imperfection is our condition, but even among false theories some are better than others. We should take a big idea seriously not only, and perhaps not even primarily, in virtue of thinking it might be The One, but rather to the extent that it holds the promise of giving us more understanding than we currently have. An exclusive focus on a theory's matching our feelings about every example we come up with also distracts us from the ultimate goal of this theorizing, which in my view is explanation. General ideas can point us to explanations that are true and illuminating in particular domains, even if as we find an idea's boundaries it must cease to be seen as generally applicable.

I see the history of 20th century theorizing about knowledge not as converging to failure but as a sequence of ideas that were developed in enough detail to brightly illuminate the subject matter by uncovering features of knowledge we had not noticed before, some by predicting them, others by tripping over them as obstacles. Alvin Goldman's causal theory of knowledge, for example, exposed the neglected role of the fact that in successful perceptual knowledge our beliefs are connected to the world in a causal and historical way (Goldman 1967). Surprisingly, causal requirements were not enough to explain why the beliefs we count as knowledge are not true accidentally, since a roll of the roulette wheel is also a causal history. However, Goldman's process reliabilism (Goldman 1979) exposed another pro-epistemic feature, reliable production of true beliefs, that both provided an explanation of why causation helps when it does, and addressed this particular accidentality problem. Similarly, Keith Lehrer and James Paxson (Lehrer and Paxson 1969) uncovered the important role of the defeasibility of evidence in our judgments of whether a belief is justified. The fact that these epistemic features had not been discussed before in traditional epistemology is sufficient to show that our understanding of knowledge has been improved, despite the fact that the causal theory, the indefeasibility theory, and even process reliabilism have important counterexamples and problems. We would understand much less than we do had it not been for the panic induced by Edmund Gettier (Gettier 1963).

One may wonder how definitions of concepts can be explanatory when they give rise to theories that aren't true. Traditionally among epistemologists the truth of a theory of the criteria for knowledge has ultimately been measured against intuitive judgment of whether a case of belief certified as knowledge by the theory really is; we ask ourselves whether we would call that case "knowledge" if we met it on the street. One problem with this standard is that there are many cases where intuitions give no clear answer; we might, and I do, conclude that the ordinary concept of knowledge is not fully determinate. There are other cases where conflicting intuitions arise for different people, or even one and the same person; the ordinary concept of knowledge probably involves a number of overlapping but not identical heuristics that yield different answers in the non-overlapping areas. The ordinary concept of knowledge is not rigorous enough or determinate enough, and possibly not even univocal enough, to support the traditional method of evaluating theories of knowledge. And I doubt that there is a unique a priori concept hiding behind the haze, at least of the sort that has traditionally been sought. We should be less confident than people often are in the weight we ask an intuition to bear. However, here, as in natural science, casual observation of cases and raw data are only part of our evidence. We also have knowledge of familiar general properties of the phenomenon we are interested in, here, knowledge, and these also have a constraining role on theorizing, as I will explain.

The pervasive existence of conflicting or absent intuitions about cases makes what it is for a theory of knowledge to be true problematic. What should it be true *to* where the data give conflicting answers, or none at all? My response is to take intuitions as defeasible constraints on theorizing, and to offer what are, quite literally, definitions, in part stipulative, for the purpose of delineating rigorous concepts that might be even more useful than the ones we encounter in ordinary language. In this way, my "What is?" questions are not Socratic, but Carnapian (Carnap 1950, 3-8). However, though I deal in definitions, the relationships between well-defined concepts are of course factual, in whatever way logical and mathematical relations are factual, and these relationships are a key part of how definitions can be the core of explanations and keys to their discovery. These definitions can interact with not only intuitions about cases, but our pre-theoretic understanding of general properties that also need explanation. If I define a simple concept like tracking, and find that its possession implies possession of another rigorously defined property that looks a lot like what Bacon described as an ability to exploit the rules that Nature herself employs (as I do – Chapters 1, 4), and if I find that what we have in what we call "knowledge" looks again and again like tracking (as I do – Chapters 1 – 4),

I will have thereby offered a plausible explanation of the truism that knowledge brings power, namely, that the thing we call "knowledge" involves tracking, and tracking brings power.

Truisms have exceptions, of course. There may be cases where something we call "knowledge" doesn't bring power, judging intuitively. In such a case, an advocate of my theory might look for some special circumstance that explained why tracking didn't bring power in these cases either. If this went well, then the case would count merely against the universality of Bacon's dictum. A different kind of exception is possible, where we have a case we are happy to call "knowledge" and which seems to give power, but where tracking is absent. This would suggest the possibility of more than one kind of power that knowledge bestows and call for an attempt to identify what simple features may be at the bottom of that. In such a case, wouldn't the tracking theory be wrong? Yes and no. The theory would have a mismatch with the data if asserted as a claim about all of it. But it would still be an explanation of knowledge's property of power in the part of the knowledge domain where the association between tracking and power exists in the instantiations of the rigorous versions of those properties. But isn't restricting the scope of a claim just a way of evading falsification? Isn't theorizing in this way a friction-free and unconstrained endeavor? Definitely not; after all, the area of mismatch would have been admitted. But also, ultimately, the evaluation of the value of a definition is in whether it plays a role in explaining something by identifying a relation between two properties that we had found instantiated together. A key point here is that having the scope of an explanation restricted does not by itself mean we have ceased to have any explanation at all.

Why should a theory with mismatches be allowed to go on to live another day? Because, or to the extent that, there remains a domain in which it does correspond to and explain a correlation between properties (not only cases), and because there may yet be appropriate correlations between the theoretical concepts, here tracking and yet other properties. Further constraints that limit evasion come from the side of what counts as a good explanation. If the domain over which the relation between tracking and the intuitive notion of power shrunk to one case, for example, the concept of tracking would be useless to explaining the power property since it would have a trivial scope of application. And if the domain of real cases that instantiate the relation between tracking and power were found to overlap only slightly with cases we intuitively call knowledge, then the process of definition would have taken us to a different subject matter from the subject of knowledge that we started with. It might be a worthy subject matter, but not the one we were trying to understand. Intuition judgments have a role in keeping us on topic.

Since the scope of applicability of a new, rigorous definition is liable to shrink under further investigation, one might question the wisdom of even attempting to give fully general simple theories. The effort is valuable, though, because greater simplicity of a theory and greater breadth of its true implications are key features of its ability to unify our knowledge of phenomena, and unification is one key to the kind of explanations I am looking for. This focus on explanation as unification is also one key to my demand for rigor. Rigor not only makes the boundaries of the defined concept definite and more efficient to evaluate in examples, but also its often-lamented abstractness and retreat from detail bring a tendency to simplicity. The fact that the tracking conditions, to be discussed below, are profoundly simple and yet highly consequential contributes to the high potential for explanation. So too for other big ideas.

A theory of knowledge should be judged, then, not just by its tally of counterexamples, but also by its promise of explanations, and especially by the range of instances of the illuminating conceptual correlations it uncovers. It is by this standard that I found the tracking theory of knowledge to have been greatly underestimated, and saw that further development of the view would be fruitful. I can only briefly discuss a selection of the issues here.

Nozick's original tracking theory of knowledge (Nozick 1981) focused on the following counterfactuals as requirements:

1) If p weren't true, then S wouldn't believe it.

2) If p were true, then S would believe it.

The appeal of these conditions was mainly seen at the time as immediate intuition—especially about 1)—and the nice trick that 1) played on skepticism, since it is possible to 1)-track that there is a table in front of us, without 1)-tracking that we are not brains in vats. We can thus concede something to the skeptic without any obligation to worry about our ordinary knowledge. Predictably, these conditions ran fairly quickly into a blinding snowdrift of counterexamples. These are chronicled in Chapters 2 and 3 of *TT*, as they are turned back, group by group, via two modifications of the original conditions. One is to reformulate the tracking conditions in terms of conditional probability:

1) $P(\text{-b(p)}/\text{-p}) > s$, where $.95 < s < 1$

2) $P(\text{b(p)}/\text{p}) > t$, where $.95 < t < 1$,

Where "b(p)" says "S believes p," and the thresholds s and t are determined by the disutilities of the corresponding errors in the subject's context. The other modification is to impose closure of knowledge under known implication. This means that it is possible to know p not only by tracking p, but also by knowing that p is implied by some q that you track. You have to track some statement to know p, but it need not be p. This relaxation of the tracking view is achieved by a recursion clause. It is imposed because I think, largely on the basis of arguments that have already been made, that knowledge is closed, and because though non-closure is a great move on skepticism, there's no reason to think it is more natural than closure is for a tracking theory. Objections to the brute force method of imposing the property of closure are not compelling: all the theories of knowledge we know of have to do this to get closure, as we see when formulations become rigorous. (Consider, e.g., the recursion clause needed in process reliabilism in Goldman 1979, and in the safety view. Deduction, though a reliable and safe process, does not preserve these properties in the conclusion. And internalists who deny closure must think the property is independent of the concept of justification.)

Just as important as the details is the much broader appeal of the tracking idea. Condition 1, *variation*, says the knower has an ability to pick up on the matters that indicate p's not holding, and to manage belief in p accordingly, while condition 2, *adherence*, says the knower has a sense of those matters that are relevant to p's holding, a sense of which are not and an ability and disposition to believe in response to the former while also ignoring the latter. The idea behind both conditions is that a knower must be responsive to the world, in the midst and process of its stabilities and variations. This shows immediately why knowledge is more valuable than mere true belief: true belief is a state at a time, which tells us nothing about whether you have a chance of staying in synch with the world over time. (See Chapter 1.) The dynamical aspect of these tracking conditions also immediately explains why knowledge would give us an advantage in the attempt to exploit the rules Nature herself uses. To the extent that we are responsive to her regularities and irregularities, we must be picking up on at least some of her signals concerning p and not-p. In order to support tracking those signals must be regular indicators of p and not-p, which means they must be kept in regular synch with nature's laws. Thus if we track, our beliefs respond in accord with laws (Chapters 1, 4).

Responsiveness is not a vacuous abstraction, for no other theory of knowledge we have takes this direction of fit—our ability to follow the world—as essential. The world's state is the independent variable and your beliefs are the dependent variables that tracking puts conditions

on. The popular property of safety—if I were to believe it, it would be true (not easily false)—involves only the opposite direction. This is inadequate, since though if I have power to make the world conform to my beliefs, that will bring safety with it, safety won't give power to me. (Multiple contrasts between safety and tracking (also known as "sensitivity") are discussed in Chapter 4.) Process reliabilism is at a less dramatic but real disadvantage here too: it says, roughly, that if I know then I used a process that will give me a true belief about p most of the time. But what if the world is such that there's a decent probability that that process I used would stop working for this purpose? We have no reassurances about that case, whereas with tracking we do, if this case is a probable enough eventuality (Chapter 3). Internalist theories of knowledge will tend to have trouble explaining its power, since they typically require no robust relationship at all between one's belief and the world. Typically the only external requirement on knowledge for an internalist view is that the belief be true, but merely true belief may be so accidentally, and that gives us no assurances about whether believing p is the best way to manage ourselves as the world moves on and we try to work on it.

The original tracking theory had a problem with knowledge of logical and mathematical truths (as does every theory of knowledge, I argue, Chapter 4). The variation condition for these truths ends up being either undefined or trivially fulfilled, since it is not possible for these statements to be false. However, the core idea behind the tracking theory is *responsiveness*, and it makes sense that to a proposition whose truth value behaves quite differently than those of empirical propositions, we owe a different kind of responsiveness. The place where appreciation of logical truths should make us believe differently in different situations is with implication. If p implies q and someone believes p, but were he to form a belief on the matter of q, that belief would be in not-q, then we would have to say that he does not know that p implies q. To know that p implies q requires having your beliefs in p and q be properly responsive *to each other*. This view also explains why it is proper to believe a logical truth itself come what may. It is not just because it will be true come what may, but because a logical truth is implied *by every proposition*. Therefore, by the account of knowledge of logical implication just described, you should be disposed to believe a logical truth r, assuming you have a belief at all, whatever your other beliefs might be. Knowledge of necessary truths is fallible—you might have been wrong—because the responsiveness you must have is not to the difference between the truth and the impossible falsity of the logically true statement, but to the relations these truths have to other propositions and to the relations they impose on the

truth-values of other propositions. The required dispositions are dispositions among your beliefs, so it is both possible for you to fail, and possible for us to count you as knowing even if your dispositions are not perfect. (See Chapter 4.)

Evidence and knowledge have a relationship that, largely because of a division of labor, never gets discussed: the better our evidence for p, the more likely we are to know p. One might think this can easily be explained by someone with a probabilistic view of evidence. After all, for her e is evidence for h if e raises the probability of h. However, that only shows that evidence makes your belief more likely to be *true*, and we've known for a long time that the truth of a belief is not enough to make it knowledge. Better evidence not only gives us a greater chance of being right, but also more of that extra robustness that knowledge has. In Chapter 5, and independently of the tracking theory of knowledge, I defend the Likelihood Ratio as the best measure of evidence, by my own and other peoples' arguments. This is a ratio of the probability of e given the truth of the hypothesis to the probability of e given its falsity. In other words, real evidence must discriminate between the truth and falsity of the hypothesis; the more it does this, the better it is. It will be better when the numerator is higher and the denominator is lower, that is, roughly, when e would allow us to better fulfill the adherence condition, and the variation condition respectively. This is the core of the argument in Chapter 5 showing a strikingly smooth connection between evidence and knowledge so understood. I go on to develop a confirmation theory that decomposes the information in evidence in a different way from standard Bayesianism, and that has advantages in allowing us to evaluate the impact of our evidence in some cases where we don't have a lot of background knowledge. It allows us to have what I call "leverage" in this and other difficult cases. (For an important correction to my application of this scheme see Barnes 2008.)

The final chapter of *TT* shows the relevance of confirmation theory to scientific realism. That is, the question what evidence is constrains our answers to the question how much knowledge of the world our evidence can and actually does get us. I argue that Bas van Fraassen's Constructive Empiricism, the best-known version of anti-realism, requires an assumption I call Equal Punishment (EP): no observational evidence can disconfirm a theory more than it disconfirms what that theory says about observables. The only plausible confirmation measure we know of that yields this result is the ratio measure (e confirms h to a higher degree the greater is the ratio $P(h/e)/P(h)$). However, the ratio measure, like all others, requires evaluation of claims about unobservables in order to ascertain whether we can legitimately make an

inference from what is observed to general claims about observables. In order to avoid radical skepticism, the Constructive Empiricist claims we can do the latter. However, the former, evaluation of unobservables, is necessary for the latter, and if he says we can do the former, he has succumbed to realism. The distinction between observable/unobservable is sensible, I grant, but it's not the issue in confirmation, and can't get the anti-realist traction for a sensible middle ground between realist and skeptical positions.

The strategy anti-realists have often taken, of making claims about the limits in principle of what our evidence can confirm seem to me unwise anyway. Philosophers have regularly been embarrassed by science in our claims that one or other thing is not possible. I argue that we have actually gotten beyond the observable line, in cases like pregnancy tests, for example. However, the folly of the realist, one who thinks our best-tested theories are approximately true, has typically been to vastly overestimate how far we've gone up the ladder of confirming general theories. Using the measure most favorable to the realist point of view, the Likelihood Ratio, we see that our high-level theories have not been confirmed, due to the difficulty of evaluating the probability of the evidence given the negation of the hypothesis (the "catch-all").

Many have thought that the problem of the catch-all is one of limits on our ability to conceive alternative theories. I argue that conceivability is irrelevant to this problem, because we now are able to evaluate large classes of theories without describing each theory or even idea, or even subclass of theories, in that class. Our methods have gotten much better over the history of science—an optimistic induction to put beside the familiar pessimistic one. So, contra the new pessimistic induction over the history of science offered by Kyle Stanford (Stanford 2006), which is based on limits of conceivability, we don't have grounds for declaring limits in principle to what new evidence may be able to confirm. The optimistic induction I proposed in the book is developed in Roush 2009.

References

Barnes, Eric Christian (2008), "Evidence and Leverage: Comment on Roush," *British Journal for the Philosophy of Science* 59: 549–557.

Carnap, Rudolf (1950), *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Gettier, Edmund (1963), "Is Justified True Belief Knowledge?" *Analysis* 23: 121–3.

Goldman, Alvin (1967), "A Causal Theory of Knowing." *Journal of Philosophy* 64: 357–72.

—— (1979), "What is Justified Belief?" in G. S. Pappas (ed.), *Justification and Knowledge*. Dordrecht: D. Reidel, 1–23.
Lehrer, Keith and Thomas Paxson, Jr. (1969), *Journal of Philosophy* 66: 225–37.
Nozick, Robert (1981), *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
Stanford, P. Kyle (2006), *Exceeding our Grasp: Science, History, and the Problem of Unconceived Alternatives*. New York: Oxford University Press.
Roush, Sherrilyn (2005), *Tracking Truth: Knowledge, Evidence, and Science*. Oxford: Oxford University Press.
—— (2009), "Optimism about the Pessimistic Induction," *New Waves in Philosophy of Science*. Palgrave MacMillan.