Testability and the Unity of Science

Author(s): Sherrilyn Roush

Source: *The Journal of Philosophy*, Nov., 2004, Vol. 101, No. 11 (Nov., 2004), pp. 555–573

Published by: Journal of Philosophy, Inc.

Stable URL: http://www.jstor.com/stable/3655667

TESTABILITY AND THE UNITY OF SCIENCE*

Several philosophers have accepted or presumed that the aim of making science unified is in conflict with the aim of keeping theories testable.[1] At issue is unity of domain as distinct from unity of method. Making science more unified in the sense intended gives rise to more comprehensive theories, and eventually to one theory so comprehensive that everything falls within the scope of the theory's claims, including any measurement apparatus used to test it. Since such a theory thus makes claims about the processes that produce the evidence used to test it, some have thought that evidence so produced does not represent a test of the theory that is independent of the theory, and thereby that either the aim of unity or the commitment to testing theories independently of themselves must be given up. Ian Hacking affirmed this tension when he opted for the first horn of the dilemma: "it is precisely the disunity of science that allows us to observe (deploying one massive batch of theoretical assumptions) another aspect of nature (about which we have an unconnected bunch of ideas)" in such a way that the observations are informative for testing.[2] I will argue that the idea that there is a tension between

[1] See Ian Hacking, *Representing and Intervening: Introductory Topics in the Philosophy of the Natural Sciences* (New York: Cambridge, 1983), p. 183; Martin Carrier, "Circles without Circularity: Testing Theories by Theory-Laden Observations," in J.R. Brown and J. Mittelstrass, eds., *An Intimate Relation: Studies in the History and Philosophy of Science* (Boston: Kluwer, 1989), pp. 405–28, esp. pp. 409–10, 423; Peter Kosso, "Science and Objectivity," this JOURNAL, LXXXVI, 5 (May 1989): 245–57, especially p. 246; and Robert G. Hudson, "Background Independence and the Causation of Observations," *Studies in History and Philosophy of Science*, XXV (1994): 595–612, especially p. 603. All of these authors accept Hacking's dilemma, discussed below, that there is a trade off between unity and testability, in some form that I will reject.

[2] *Representing and Intervening*, p. 183.

these two aims is largely based on a conflation of factual and epistemic independence, and that although the de facto disunity of today's sciences does make it easy to avoid certain testability problems, disunity is not necessary to maintaining testability of theories. Thus, the aim to preserve the independent testability of theories is not a reason to give up the goal of unifying scientific theories.

To give the view I oppose the greatest chance of being right I will understand the goal of unifying science in a strong sense: I will suppose that the aim is to find one (physical) theory that explains everything. I will not try to decide whether this means only that tokens must each find some correlate or other in the theory, or also that types in upper-level theories must be reduced to types in the theory. However, I will suppose that the fact that the theory sought explains everything means, in particular, that it gives an account of why certain (possibly upper-level) empirical processes are or are not reliable indicators for certain (possibly lower-level) states, for every such process and state.[3] The theory's explaining everything will mean, at least, that the theory is probabilistically relevant to every statement about the world—like the know-it-all, such a theory says something about everything.[4] This latter assumption will be enough to generate the apparent problem.

In particular, the theory will be probabilistically relevant to any statement as to whether a given process is a reliable indicator of a given state. To fix ideas, let us say that a process whose outcome will be either $+$ or $-$ is a reliable indicator of a given state, S, if and only if the process is such that $P(S/+)$ is high and $P(S/-)$ is low.[5] The assumption I make, then, is that the truth or falsity of the know-it-all theory makes a difference to these probabilities. If so, then because the know-it-all theory predicts either S or not-S when $+$ or $-$ is the outcome respectively (since it predicts something about everything),

---

[3] I am using the terms 'upper-level' and 'lower-level' as in discussions of reductionism, where reduced theories are upper-level and their entities are often composed of the entities of the lower-level reducing theory. Physics is lower level, psychology upper level.

[4] As Wesley Salmon argued in *Scientific Explanation and the Causal Structure of the World* (Princeton: University Press, 1984), pp. 84–89, 46, there are cases of explanation in which the explanans lowers the probability of the explanandum, and it is even typical for part of the explanans to do so. This is not a problem for our assumption since negative probabilistic relevance is a kind of probabilistic relevance. It is anyway clear that negative probabilistic relevance of a theory to test procedures has as much potential as positive relevance does to allow the theory to affect the assessment of a test of it, since through negative relevance to the reliability of a test procedure the theory may veto the credibility of an outcome that stands ready to falsify it.

[5] The expression '$P(S/+)$' should be read "the probability of S given $+$," and $P(S/-)$ should be read "the probability of S given $-$."

it is clear that this theory can affect our assessment of whether the fact that, say, + occurred supports, undermines, or does not affect our right to believe this same theory. It would do so by having a view, so to speak, about whether the occurrence of + does or does not indicate that S obtains. Thus, for us to take a view about whether S obtains is to agree or disagree with the theory under test in deciding whether the evidence counts for or against that theory.

I will suppose in addition that the theory sought is internally unified, that is, that the accounts it gives of empirical processes that might be used to test it could not easily be separated from the accounts it gives of other things, perhaps because it has an ontology with only one or a few types of entity, or it has a single set of laws. At a minimum, the theory is not tacked together. This is a property of theories that has actually been sought by physicists. For example, Albert Einstein demanded of physical theories that they give an account of all measuring devices whose results are relevant to their testing in a way that does not make those devices a distinct class of entity from the other entities that the theory posits, and faulted special relativity for not achieving this aim.[6]

Internal unity of the sought-for theory makes the task of this paper more difficult because it is a property that will tend to block an easy way of getting out of problems that unity might present for testability: it will prevent us from testing the theory by testing its pieces one at a time. Taking this property on board makes the problem about testability look especially acute in light of typical discussions of independence of evidence from the theory under test. A natural way to analyze how far a theory's being used to give an account of observational evidence is independent of the theory under test, when the two are the same theory, is to divide that theory into participant and beneficiary parts and look at how far the truth conditions of those two parts are independent, as Peter Kosso does.[7] The internal unity that I have just attributed to the maximally unified theory suggests that parts of the theory under test will be very difficult to separate in this way. If so, then on Kosso's account it would follow immediately that testability of that theory independently of itself is close to impossible.

One might protest against the ethereal quality of posing the question about unity in the way I have, about a maximally unified theory, since we do not seem very close to finding such a thing. However,

---

[6] "Autobiographical Notes," in P.A. Schilpp, ed., *Albert Einstein: Philosopher-Scientist* (La Salle: Open Court, 1949), pp. 2–95, here p. 59.

[7] "Science and Objectivity," pp. 252, 257.

physicists do seek a grand unified theory that would account for every
scale of the universe, and a significant number of people do believe,
rightly or wrongly, that such a theory would provide an explanation
of everything, at every level. One might think that the kind of thing
physicists seek would not qualify as maximally unified because there
would always be separate boundary conditions, but some physicists
even take it as an ideal goal for physical theory that all brute numbers
eventually be predicted from laws, however unattainable that might
seem. If the most unified theory we ever found did not make predic-
tions unless it was combined with quite independent boundary condi-
tions, the question I am asking here could still easily arise for the
theory because that highly unified theory would likely be probabilis-
tically relevant to auxiliaries assumed about measuring devices used to
test it. Asking the question in the manner I do in this paper is simpler
than carrying along a discussion of boundary conditions, and because the
results defend unity, they defend those slightly less unified cases as well.

   If the disunitarians were right, then that would be significant. It
would mean that scientists should give up on finding a certain kind
of theory, because even if they succeeded in finding the type of thing
they are looking for, it would not be possible to test that theory
empirically. An untestable theory might be good speculative metaphys-
ics, but it would not be good physics. Moreover, the positive relevance
view of what evidence is, which is popular among philosophers of
science, where e is evidence for H just in case e is positively probabi-
listically relevant to H, that is, just in case $P(H/e) > P(H)$, appears to
add further plausibility to the disunitarian claim by treating judgments
about reliability of the evidence and confirmation of the theory in
commensurable formats. If what it takes to be evidence for something
is raising the probability of that thing, and the most unified theory
is probabilistically relevant to everything, then it appears that that
theory can raise (or lower) the probability of claims that will then
raise (or lower) *its* probability, namely the claims about whether the
test process was reliable. Circular justification appears to threaten.

   In the popular Bayesian conception of scientific reasoning, hypoth-
eses, and all other statements, are assigned probabilities, and on many
versions of the Bayesian view probabilities translate fairly directly into
degrees of belief, smoothing further the path to an apparent problem.
In particular, a prior probability must be assigned to the unified
theory; it is not an option to imagine oneself as withholding judgment
on the theory until the evidence is assessed, because without a prior
probability the theory can have no posterior probability. The force
of that prior probability propagates over all of the statements the
theory is relevant to—if the theory is probabilistically relevant to a

matter, then the prior probability of the theory will affect the probabil-
ity that matter gets. In the case we are imagining, the prior probability
of the theory will apparently have an effect on the probabilities of
the auxiliaries used to decide whether a given test outcome should
be counted for or against that very theory.

Thus the question I am addressing here is more timely and pressing
than it may at first appear. The arguments in this paper aim for a
strong conclusion insofar as they respond to the disunitarian idea
without denying the positive relevance conception of evidence, the
Bayesian conception just described, or the requirement of indepen-
dent evidence for auxiliaries.[8] Indeed, I will show that, contrary to
appearances, lack of probabilistic relevance between theory and relia-
bility-auxiliary is well suited to measuring epistemic independence of
an account of evidence from the theory under test.

Even so, one may say, this epistemological argument against the
unity of science was never as influential as other arguments for the
same conclusion. Since, one may think, the disunitarians have won
those other arguments, turning back this argument would not get us
very far. However, this would be a mistaken impression of the situation;
in fact, this epistemological argument is stronger than typical argu-
ments against reductionism, because it applies even to reductionism's
weaker cousin. One of the strongest arguments against the domain
unity of science was Jerry Fodor's argument that the plausible doctrine
of token physicalism—every event is a physical event—is distinct from
and significantly weaker than the less plausible unitarian thesis of
reductionism, which implies that every kind corresponds to a physical
kind.[9] Unitarians can do all they need to do with the former, he
argued, and must not claim the latter on the basis of the plausibility
of the former. Like Fodor, one might think that, despite all of the

---

[8] I assume that what is wrong with judgments of the evidence lacking independence
from the theory under test is circularity in reasoning, by which I mean use of a
theory as a reason to believe or discount evidence for or against that theory, and
that such circularity is a bad thing. For a defense of this, see my "Testability and
Candor," forthcoming in *Synthese*. Clark Glymour's bootstrapping account of confir-
mation in *Theory and Evidence* (Princeton: University Press, 1980) is well suited to
understanding the relativity of evidence to the theory under test, but is not concerned
enough about independence of evidence for auxiliaries—see *Theory and Evidence*,
pp. 114–21. His condition that "to test a hypothesis we must do something that could
result in presumptive evidence against the hypothesis" is necessary but insufficient
for independence of evidence from the theory under test, for reasons developed in
"Testability and Candor." Thus, I am here classifying Glymour's view as giving up
on the requirement of independence of evidence from the theory under test.

[9] "Special Sciences, or The Disunity of Science as a Working Hypothesis," *Synthese*,
xxviii (1974): 97–115.

arguments against the domain unity of science, a retreat to token physi-calism is still a safe bet. The epistemological argument under discussion threatens to destroy even this safe retreat, since a physical theory that made good on the token physicalist claim would say something about every event, and thus would be probabilistically relevant to every event, including all of those involved in the setups used to test it.

We must ask if such a theory could be tested, and if the answer is "no" then even token physicalism cannot be recommended as a guide to the development of scientific theory. In that case it could still be that token physicalism is true, but no theory that exemplified it could be considered genuinely testable, and hence really scientific. Disuni-tarians and unitarians alike seem to regard token physicalism as innoc-uous. Ironically, if the epistemic disunitarian argument is right then this doctrine may be true but pernicious as a recommendation for what kind of physical theory to look for. It would be wrong even to look for a comprehensive, unified theory within physics construed as a special science.

My procedure here will be to imagine that we have a maximally unified theory—call it "Behemoth"—and investigate what it would be like to test it. I will consider what is required for independence of evidence and argue that Behemoth's scope does not pose an un-answerable challenge to finding it. If unity is a problem for testability, then Behemoth will be the hardest possible case, so dealing with it will be sufficient to defend unity. The potential source of problems is not hard to see. In the course of evaluating evidence from a test as to whether it tells for or against a theory, or reveals nothing at all, one must evaluate the reliability of the test processes as indicators of theoretical quantities. If they are deemed unreliable then we have reason to discount the outcomes of the test, whether they were unfa-vorable or favorable to the theory. Such discounting will favor or fail to favor the theory respectively. We would not want to believe mistakenly that they were unreliable, lest we throw out valuable con-firming or disconfirming evidence about the theory. But if they are unreliable, then it is important that we believe they are because other-wise we might count results as favoring or disfavoring a theory when they really reveal nothing at all. In other words, it is important for testing the theory that our judgments of the reliability of the process producing the evidence are correct; they also should not be prejudiced by that theory.

As I have said, unifying science would make one theory that had something to say about the reliability of every process that could be used to test it. That is, considering the matters in isolation, the theory would be probabilistically relevant to whether S is the case given $+$

and whether S is the case given $-$ for every state S, and every set of possible outcomes of a measuring process, '+' and '$-$'.[10] For every S, that S obtains will be confirming or disconfirming of the theory (to some extent)—because the theory will have predicted either S or not-S—and the theory will apparently have the power to affect whether we have a right to believe that S obtains. Circularity—the theory's passing judgment on itself—appears to threaten.

THE NORMAL CASE

To judge whether these suggestions hold up, we need to consider the matter step by step, starting with the normal case. In many actual cases in our current un-unified science, what the reliability judgments in question are about has nothing to do with the subject matter of the theory under scrutiny. In studying cells under a microscope, for example, the reliability of our evidence will depend to a large extent on the workings of that instrument. But the optical processes transpiring in a microscope do not fall within the scope of any hypothesis about cells. No knowledge of cells is required to know something about the workings of a microscope. Nor would knowledge of cells, if we had it, help us to know about the microscope. This is in the first place because cells and lenses are factually, and therefore probabilistically, irrelevant to each other; cell claims (C) being true would make no difference to whether lens claims (L) were true. (Therefore, $P(L/C) = P(L)$ whether probability is construed objectively or subjectively.[11]) The intuition that says evidence, to be evidence, should be probabilistically relevant to the hypothesis it is evidence for, is embodied in the Bayesian predilection for counting e as evidence for H only when $P(H/e) > P(H)$, which holds just in case $P(e/H) > P(e)$.[12]

The hypothesis about cells that is under test by a microscope could not possibly be evidence for or against statements of the reliability of the microscope. No (rational) use could be made of any statement that

---

[10] Obviously, in general a measuring process may have any number of possible outcomes. I persist in using '+' and '$-$' for the possible outcomes to represent the division between those that are favorable and those that are unfavorable to the theory under test. For simplicity here, I am assuming that outcomes can be judged favorable or unfavorable independently of assumptions about the truth of the theory under test.

[11] This should be read: "The probability of L given C is equal to the probability of L."

[12] Note that probabilistic relevance is a weaker condition, easier to fulfill, than logical relevance, and probabilistic independence is therefore a stronger condition than logical independence. Two statements are logically independent if the truth value of neither fixes the truth value of the other, but probabilistically independent only if the truth value of one makes no difference whatever to the probability of the other.

is a deductive consequence of the hypothesis in the assessment of the reliability of the microscope, because that hypothesis has nothing to offer.[13] This example illustrates why probabilistic relevance is necessary between two things if one is to be evidence for the other. Those worried about the consequences of unity for testability are right to think that probabilistic irrelevance (independence) between theory under test and the account of the measuring apparatus is a salutary feature of ordinary (so far un-unified) science. Our beliefs or prejudices about the cell hypothesis we are testing have no rational route through which to contaminate our beliefs about the reliability of the instrument used to test it, the microscope.

<div align="center">PROBABILISTIC REVELANCE AND EVIDENCE</div>

If there were probabilistic relevance between hypothesis under test and process happening in the measuring instrument, considered in isolation, then this easy insurance that our knowledge of the latter was independent of our beliefs and prejudices about the former would be gone. It is tempting to suppose that this automatically means trouble in the form of circularity in our testing. After all, in this case whether or not the theory is true makes a difference to the probability that the process used in testing (that theory) was reliable.

It does follow that there is a problem on Kosso's account of independence of evidence from the theory under test, mentioned earlier. This is because his recipe for determining whether evidence is independent from the theory under test is to ask which theoretical statements we accept in the account of the measuring apparatus, and then to ask whether those statements are independent in their truth conditions from the statements of the theory under test. If there is an identity between any statements in the two sets, as there will be if there is probabilistic relevance between theory under test and theory of the instrument (whether considered in isolation or all things considered), then on this view we automatically get lack of epistemic independence between the two.[14]

The unitarian must hope that there is a glitch both in the intuitive idea, and in Kosso's analysis. But as for the intuitive idea, if we take,

---

[13] In other words, if we let our prejudices about cells affect the probabilities we assigned to claims about the microscope, we would be irrational. The axioms of probability do a great service here in disqualifying as irrational all manner of influence by irrelevant factors. Of course, the axioms have no remedy against a subject who assigns probabilities in such a way as to make one thing relevant to another when it is not in fact relevant. To disqualify that requires further constraints on the probability function.

[14] "Science and Objectivity," pp. 252, 257.

as the Bayesians do, positive probabilistic relevance as the defining property of evidence, then if there is positive probabilistic relevance of the theory under test to the account of the instrument, then the theory under test is evidence for a certain account of the instrument, which account will be crucial in determining whether the results of the experiment favor or undermine the theory. To defend unity, then, we have to ask whether there is any good reason to deny the positive relevance view of evidence, or to think we are mistaken about its implications.

The obvious place to put pressure is on the assumption that positive probabilistic relevance of A to B is sufficient for A to be evidence for B. And there is intuitive reason for doing so. It is plausible, intuitively, that a further condition for A to be evidence for B, beyond A's relevance to B, is that A is one of our reasons for believing B, part of the basis for our belief in B. This suggests that it is wrong to think that relevance of A to B is generally sufficient for A's being evidence for B, because we can imagine having a belief that is relevant to another belief but which is not the basis on which we hold that other belief; we may have other sufficient reasons for the second belief. There are cases where the relevance of A to B is sufficient for A to be evidence for B if anything is, but these are cases where A is the only belief we have that is relevant to B.

That a belief may not be the basis for another of our beliefs even if the first is relevant to the second—in the sense that whether the first is true makes a difference to whether the second is true—is the intuitive key to the primary way in which theories or hypotheses probabilistically relevant to the reliability of the processes used to produce evidence for or against them can fail to result in circular testing. This is also the area in which Kosso's analysis of independence of evidence fails: though Kosso was rightly aiming at an epistemic notion of independence, the particular mechanics of his analysis missed the mark. Because on his view we first ask which statements about the instrument are accepted and then ask whether those statements are independent in their truth conditions from the theory under test, we lose any purchase on whether the reasons for which we accepted the statements that make up the account of the instrument are independent of belief in the theory under test.

Kosso's analysis of independence of evidence from the theory under test does not allow for the possibility that we believe and use a claim that is a part of the theory under test without that theory being our basis for believing the claim. Yet, surely this is possible, and it is the place to look for how to turn back the disunitarian argument, as we will do in following sections. It is a separate question whether the

notion of probabilistic relevance respects this intuitive idea that one
may have a belief that is factually relevant to another, yet not one's
basis for belief in the other, but we will see that it does if we have a
proper understanding of probabilistic relevance.

THE ABNORMAL CASE

Consider a simple case in which a hypothesis is relevant to reliability
auxiliaries used to test it.[15] Our hypothesis, H, is that all fluids expand
on heating. We will test it by heating many fluids and measuring their
volumes as we take their temperatures with a mercury thermometer.
One auxiliary hypothesis, call it "X", that we want to assume about the
reliability of the instrument is obvious: mercury expands on increase of
its temperature. But assuming as we can from background knowledge
that mercury is a fluid, the hypothesis under test, H, is probabilistically
relevant to this auxiliary X; in fact, the latter is an instance of the former.

  Despite this probabilistic relevance between hypothesis under test
and auxiliary, we can have evidence for this auxiliary independently
of the hypothesis. Notice that there are other methods of measuring
temperature than glass bulb thermometers, and we can use one of
those methods, say electrical resistance, to calibrate our glass bulb
thermometer.[16] That is, our other method of measuring temperature
will tell us that the mercury's expansion does indicate a rise in temper-
ature. Let e be the claim that the electrical resistance method says
that our glass bulb thermometer accurately indicates elevation of
temperature (in a given range).

  Let us assume, just for the moment, that we know that the electrical
resistance method is reliable. Then we could use e as evidence for X,
the auxiliary, because e is relevant to X. That is,

  (1) $P(X/e) > P(X)$

But under our assumption that the electrical resistance method is
reliable, e is also (positively) probabilistically relevant to H. Another
way of saying this is that e is not probabilistically independent of

---

  [15] Allen Franklin et al. cite this example as falsifying the thesis that theory-laden
observations cannot test theories in "Can a Theory-Laden Observation Test the
Theory?" *British Journal for the Philosophy of Science*, XL (1989): 229–31. However, no
analysis is given there of why this is possible, in this case or in general.
  [16] We could also heat a closed container of mercury on a Bunsen burner and
confirm that the upper surface of the mercury rises by watching. The point is that
other processes exist for calibrating the mercury thermometer. I have chosen a
sophisticated method to avoid the suggestion that what matters is watching things
we have more direct access to than we do to the workings of the measuring instrument
we began with.

H, that is, the evidence for the auxiliary X is not probabilistically independent of the hypothesis under test. However, notice that despite this, and despite the fact that H is probabilistically relevant to X, the relevance between e and X is unaffected by H in the following sense:

$$(2) \qquad\qquad P(e/X.H) = P(e/X)$$

Since of the many instances of fluid and temperature H speaks of, e commented only on those pertaining to our glass bulb thermometer, no part of H except X—H's restriction to mercury—is relevant to e. X thus screens off the relevance of H to e.

The phenomenon of screening off is key to understanding why probabilistic relevance can respect our intuitive notion of independent evidence and the fact that it is not identical to lack of intuitive relevance or to what I have called "factual" independence. *C screens off A from B* if and only if $P(B/C) = P(B/C.A)$. In such situations, though A may be probabilistically relevant to B when the two are considered in isolation, and though C does not change that fact, when C screens off A from B, assuming C true renders that relation between B and A, as it were, impotent.[17] Moreover, if C screens off A from B, then according to a probability function that assigns probability 1 to C (effectively taking C to be true), the relevance between B and A is not merely impotent but nonexistent, since B and A are *not probabilistically relevant*. In our example, e and H are no less intuitively relevant to each other, and are no more factually independent of each other, if we suppose that X is true, but if X is taken into account then e and H are probabilistically independent.

That condition (2) is fulfilled is not enough, though, for the kind of independence we need in order to avoid circularity of testing, for two reasons. Although, as in (1), e raises the probability of X when the two are considered in isolation, it also appears that H screens off the relevance of e to X, that is, that e is not evidence for X if H is assumed. Second, we want not only that e can affect the probability of X, but also that our grounds for believing e itself are "independent" of our beliefs about H. By this I mean that our grounds for believing e are secure or securable whatever our beliefs about the truth or falsity of H. It is only this kind of independence that will prevent a circularity in which our belief in H is part of our grounds for believing e, which is

[17] Because if C screens off A from B then it follows that C screens off B from A, when C screens off A from B we can say "C screens off the relevance between A and B."

part of our grounds for deciding what to believe about H. Fortunately, addressing the second concern will also address the first.

Our grounds for believing e as I have stated it—"the electrical resistance method says that our glass bulb thermometer accurately indicates elevation of temperature (in a given range)"—can be independent of our beliefs about H, because we can verify what the electrical resistance method *says* about the temperature of the thermometer without recourse to assumptions about heating fluids. However, this is too easy because in that statement of e I left the reliability of electrical resistance as a method of measuring temperature out of consideration, assuming it as background. e will not help us to verify X, of course, unless we have some assurance that this other method is reliable. So, the main kind of independence we want amounts to this in the present case: we want our grounds for believing E, that electrical resistance is a reliable method of measuring temperature, to be securable regardless of what we believe about the truth or falsity of H.

It seems right that we can know whether the electrical resistance method is reliable regardless of what we believe about H. This is because the electrical resistance method makes no essential use of fluids or volumes. If so, then it is because H is not factually or intuitively relevant to E, which is sufficient to imply that H is not probabilistically relevant to E when H and E are considered in isolation. That is,

$$(3) \qquad\qquad P(E/H) = P(E)$$

If we are able to assign a high probability to E, it will not be because of the probability we assigned to H. Because H does not meet the first necessary condition mentioned above for being evidence for E—probabilistic relevance—H cannot be evidence for E, and so does not thereby act as evidence for e which is evidence for the auxiliary X that we need in order to test H.

This shows that H is not evidence for e, but not yet that H does not act as evidence for the auxiliary X. For this we need to evaluate whether e and E screen off H from X, that is, whether $P(X/e.E.H)$ is equal to $P(X/e.E)$. In fact, these terms will never be exactly equal, but the reason need not concern us; it is because of the extreme relevance of H to X—given that mercury is a fluid, H implies X—and no finite amount of inductive evidence can replace deductive warrant. However, e and E do act in the direction of screening off H from X, that is, E reduces the probabilistic relevance of H to X (when e is confirmed). We might say that statements like e and E screen off H from X by degrees, and to the extent that they do so H is not probabilistically relevant to X and thus not evidence for X.

We avoid an evidential circle if we use the electrical resistance method

to validate the mercury thermometer, because we avoid a probabilistic relevance circle. This shows how despite probabilistic relevance of H to X, that is, of a hypothesis to an auxiliary about the reliability of the processes used to test it, when the two are considered in isolation, we need not have a testing circularity in which H is a reason to believe evidence for itself: we can have reasons to believe X that we have grounds for regardless of our beliefs about H. The probability calculus conforms to this idea, since the statements about the results and reliability of the electrical resistance method will screen off (at least by degrees) the relevance of H to e and the relevance of H to X.

This lack of connection, however, between H's being probabilistically relevant to X, when the two are considered in isolation, and H's being evidence for X is achieved through the existence of a process for verifying X that H is not probabilistically relevant to even when the two are considered in isolation. This reaffirms that the easiest way to ensure that one thing is not used as evidence for another is for the first to fail to be factually relevant to the second, or, as here, relevant to our evidence for the second. Much as we have learned here about the difference between factual or intuitive relevance and probabilistic relevance, this kind of case cannot help us decide what to think about unity and testability. This is because there is a salient difference between H and Behemoth, the maximally unified theory: for any process used to test Behemoth—all of which Behemoth will be relevant to—there will be no other process with which to calibrate its reliability that Behemoth is not *also* relevant to, since there is no process at all that Behemoth is not relevant to. We are back again to the bare question whether the probabilistic relevance of a theory to a reliability auxiliary makes the theory evidence for the auxiliary, and how there could be evidence for the auxiliary that is independent of a theory with such reach.

THE FOUNDATIONALISM OF BASIC BELIEFS ACCESSED DIRECTLY

Where should we look for a basis for beliefs about measurement processes that is independent of the theory under test? We might think we could get leverage from the following thought: generalizations are intuitively, factually, and probabilistically relevant to their instances, yet surely their instances serve as evidence for them, and as evidence that is independent of them. That is, in the most basic case of empirical inquiry, our grounds for believing instances are what they are regardless of our beliefs about the generalizations they confirm. We can find out whether a particular swan is white on grounds that we have access to regardless of our beliefs (or prejudices) about whether all swans are white, can we not? We may think so in simple cases. We know what

counts as a swan, and what counts as white by inspection and defini-
tions. And we know that our perceptual systems are reliable, when
we do, without appeal to the generalization that all swans are white.
Any prejudice we might have about the truth value of the claim that
all swans are white does not need to play a role in determining whether
a given case is a confirming or disconfirming instance.

To suppose, however, that this is a model for how we should think
of our question about testing maximally unified theories would involve
stronger foundationalist assumptions than many might be comfort-
able with. After all, in the swan case the hypothesis under test does
not say anything about what counts as white or what counts as a swan;
it is probabilistically irrelevant to these questions. It is otherwise with
Behemoth, the theory that says something about everything. By anal-
ogy, it would say something about the equivalent of what counts as a
swan and what counts as white, because its statements say something
relevant to whether a given outcome counts as reliable. To suppose
in general, without further ado, that a theory that is relevant in this
way to the means of verifying instances has no role as a ground for the
instances seems to require supposing that there is a class of things that
are known by some basic means, like inspection, and for the grounds
of belief in which no abstract theory or consideration is required.

It is interesting, though not too surprising, that one of the resources
of foundationalism would seem to be to block the association of
greater unification with failure of independent testability, but we have
to wonder whether something similar could be achieved without the
foundationalist assumption I cited. Foundationalism's lack of popular-
ity is not the only reason to wonder: consider that Behemoth, which
is probabilistically relevant to every process, ought also to be relevant
to whatever process gives us the right to our basic beliefs. Merely citing
the existence of an immediate evidence-giving process like inspection,
though it may show why use of Behemoth as evidence for claims
designated basic is not necessary, does not articulate why Behemoth
could not or should not be competing evidence for or against those
claims. While the foundationalism I have sketched seems to me too
much to swallow with its privileging of beliefs formed by inspection,
the more pressing problem is that it is not strong enough to answer
our question.

THE SOLUTION

It is a commonplace of "new" philosophy of science that in the history
and practice of science theories are not tested alone, as if against
their negations, but are always compared to a particular rival. This
idea can help us here. That a theory wins against a particular rival in

a test is obviously insufficient to show that the theory is true, but let us set aside for the moment worries that being tested against rivals is too low an epistemic standard. It seems far fetched to suppose that we will come up with two or more Behemoth-sized candidate theories when it is so difficult to come up with even one (unified) theory that could purport to explain everything, but we will be rewarded if we indulge the fantasy for the sake of argument.

So, suppose we have two Behemoth-sized theories, each of which offers some explanation for everything, and is probabilistically relevant to everything. If our task in a test is to choose between these rivals, then, I claim, it is possible, even likely, that we can have evidence for the reliability of the test processes, and therefore for tilting in the direction of one or the other theory, independently of the theories. This is so despite the probabilistic relevance of both of them to every test process. The reason, schematically, is that it is possible, even likely, for the two rivals to agree on whether a given test process was reliable, while disagreeing in their predictions about the outcomes in the test. In such a case, the reason that the evidence we get from the test would be independent of our beliefs about which theory is better would be not that the theories are irrelevant to the test process, or that certain things are known by inspection and not in any other way, but that whichever theory you favor will give you the same view of the reliability of the test process. Your favoritism toward one or the other theory will make no difference to your view of the test process.

For example, it is certainly possible, for reasons of logic, for two theories to have different views about whether all fluids expand on heating—the matter at issue in our earlier test—while having the same view of whether mercury expands on heating under normal conditions in the medium range. If two such theories were the rivals you were considering, and the view of both was that mercury does expand on heating in the usual circumstances, then you could assume this regardless of your stand on the issue the two theories disagree about. That two rival theories have a view about every process does not imply that they have different views of every process at every level of description. In fact, given the types of theories they will be, Behemoth-sized theories are likely to agree about quite a lot. This is because it is a condition of adequacy for such a theory that it does well at predicting the things we take ourselves to know already—it should, ideally, make predictions that are the same as or better than those of all of our best less general theories in every domain, at every level where those other theories have been tested. Where two Behemoth theories will differ is in the deeper accounts they give of all such things—Is it strings beneath it all?, for example.

For present purposes we do not need to take a stand on what it is about the structure of knowledge that makes some claims more justi- fied than others, and therefore more likely to be agreed on by rival theories of everything. And we do not need to say here what our grounds are for believing these agreed-upon claims, though in my view it has to do with the claims having survived scientific scrutiny for long periods of time rather than with the claims being matters we have direct access to. It is enough for present purposes to notice that there will be quite a few such claims, and they will not all be direct observations, an advantage of this strategy over the earlier foundationalist reply.

For example, the predictions and even some of the regularities among theoretical quantities of Newton's laws of motion and of classical electro- magnetism, in their restricted ranges of applicability, will be among the things we are more justified in believing than we are either Behemoth candidate. It is important that the agreed-upon claims not all be merely observational because in many cases knowing the reliability of a test procedure requires knowing that the procedure is a trustworthy indi- cator of some theoretical quantity. Observations alone cannot provide links between observations and unobserved quantities, and those links are essential to testing theoretical claims. There is the additional fact, in favor of my overall point, that we tend to use as probes processes we understand better than the processes we are investigating; this makes it likely again that the processes used as probes will be matters the Behemoth-sized rivals agree about.

Moreover, we can throw away the ladder in this argument: since what is doing all the work is the claim that some of our claims are more justified than others, we do not need to suppose that a Behemoth has an articulated Behemoth-sized rival to suppose that it can be tested. If I am right that some of our claims are more justified than others, and that the things that need to be assumed in testing are often among the claims that are better justified, then those who favor the Behemoth theory and those who oppose it without a rival theory to offer will tend to agree on quite a lot about what happens when it gets tested, and the basis for that agreement will be already accumu- lated justified belief. "New" philosophy of science was not necessary, after all. Of course, nothing here says that our justified, or somewhat justified, beliefs about what goes on in processes we think we already understand are true. However, the question at issue here was not their truth but rather on what basis they were believed. The saving bit is that they were believed on the basis of other things than commit- ment to the theory under test.

Essentially, this argument that unity need not interfere with the testing of a Behemoth corresponds to the following fact about probabi-

listic relevance: even if for any X, the theory T is probabilistically relevant to X when T and X are considered in isolation, if there exists a C such that C screens off T from X, that is, $P(X/C.T) = P(X/C)$, then if C is in the background knowledge the probability assigned to T will not affect the probability assigned to X. The Behemoth theory may be probabilistically relevant to everything when this theory is considered in isolation from background beliefs, but it will not "govern" its own testing to the extent that background assumptions—secured antecedently and effectively given probability 1 by the probability function—screen off the relevance of Behemoth to particular processes we already understand well.

This is because the screening off of Behemoth by background knowledge means that in the testing situation Behemoth is not probabilistically relevant to the test procedures. This corresponds to the intuitive idea of a belief failing to be the basis of another belief despite its being "relevant" to that second belief. That a theory is probabilistically relevant to X when the two are considered in isolation does not imply that the theory is probabilistically relevant to X in the testing situation, and so does not imply dire consequences about the circularity of the testing. In many situations in science screening off is achieved through use of a process that is factually independent of the theory under test, as in the validation of the mercury thermometer above. But screening off can also occur simply because what screens off the hypothesis under test is part of the background knowledge. How did it get to be part of the background knowledge? The concession may be unavoidable that the testing of a Behemoth must be preceded by a great deal of un-unified empirical science in order that this background knowledge be secured with an appropriate independence from the Behemoth. However, this is not much of a concession, since we were not asking whether Thales's theory that all is water was empirically testable when he announced it, but whether we who have inherited a mountain of results of un-unified empirical science can afford to have more unified theories in the future.

Of course, none of this means that there will never be processes used in tests whose reliability as indicators of theoretical quantities the two theories under test, or the proponents and opponents of one theory, disagree about, but such theories will be testable if there will be, as it seems there often will, plenty of tests that do not have this property. To the credit of the disunitarian's argument, I do not see how any general guarantees can be made that testability can always be preserved in the face of increasing unity, and their intuition identifies the reason why. However, the challenge they have identified does not show that we must fail.

This is true even in cases where background knowledge is insuffi-
cient to screen off the relevance of theories to processes used to test
them. It has in fact happened that one of the only crucial tests we
knew how to make between two very general theories involved a
process that the rival theories disagreed about and about which
agreed-upon assumptions could not decide, but through ingenuity
and luck epistemic folly was avoided. I refer to the fact that in the
Michelson-Morley experiment one must measure the lengths of the
arms of the apparatus in order to know whether the seeing or not of
interference fringes means anything. The experimenters used a ruler
for this purpose, but even more sophisticated means of measuring
would run into the following stumbling block: the rival to the ether
theory implied that due. to Lorentz contraction that ruler (or other
measuring device) would have different lengths in the two directions![18]

Epistemic folly was avoided here because in fact the amount of
error that could be introduced by the Lorentz contraction of the
rulers could contribute only at the third or fourth order, as both H.A.
Lorentz and L. Silberstein argued, whereas Michelson and Morley
were probing second order effects. In other words, though both theo-
ries, taken in isolation, were probabilistically relevant to the disputed
assumption (in opposite directions), and this relevance could not be
screened off by background knowledge, neither was probabilistically
relevant to the crucial measurement, because of the way the apparatus
worked and a consequent mismatch between the order of the effects
probed and the order of the effects of the disputed assumption.

The idea of background assumptions we have an antecedent right
to, and their damping effect on probabilistic relevance, provides a
general recipe for seeing how we secure independent testability in
most cases where it looks like there is too much relevance for comfort:
the background assumptions screen off the probabilistic relevance of
the theory under test. For the cases where background knowledge
does not provide enough to screen off relevance, what we can say in
general is less informative, yet not trivial. The case I have just described
is exceptional in science for the combination of high ontological
overlap between the domain of the tested theory and claims about
how the measuring apparatus worked, with high generality of the
theory under test: anything that could act as a ruler would have had
the same problem. But the fact that such cases are exceptional in

    [18] Ronald Laymon, "The Michelson-Morley Experiment and the Appraisal of Theo-
ries," in Arthur Donovan et al., eds., *Scrutinizing Science: Empirical Studies of Scientific
Change* (Baltimore: Johns Hopkins, 1992), pp. 245–66.

actual science does not get us off the hook: these are precisely the kinds of situations that increasing unity makes more likely.

Though this case shows that there are situations where the screening off solution outlined above does not work, however, it also shows that other types of solution are possible. And the two types of solution yield a general lesson we can take from the preceding: to preserve testability when a theory under test is probabilistically relevant to the test process when the two are considered in isolation, find a way to eliminate that probabilistic relevance. Screening off is one way of doing this. Setting up the experiment in such a way that the measurement is at the wrong level of generality or precision for the theory to make a difference to the probability that the measurement is reliable is another.

CONCLUSION

In this paper I have assumed a picture of confirmation as the raising of probabilities, thus giving a central place to probabilistic relevance in the conception of evidence and giving the greatest possible chances to the idea that the probabilistic relevance of a theory to everything will interfere with independence of evidence for auxiliaries used to test it. Even so, careful scrutiny of the intuitive and probabilistic notions surrounding independent evidence has shown that fears that unification of science will interfere with testability of ever more unified theories are largely unfounded: background knowledge often screens off the relevance of the theory under test to the test procedures. To the credit of the disunitarian's argument, I see no way of giving a general guarantee that independent evidence can be found in every case of testing a maximally unified theory, and that argument correctly identifies an important reason why. However, that challenge does not imply that we must fail, and the kind of ad hoc ingenuity that is required to overcome such a problem when background knowledge is not sufficient is not unlike what a scientist must bring to the other kinds of challenges involved in designing a telling experiment. Moreover, there is general advice that we can give: find a way to eliminate the probabilistic relevance of the theory under test to the specific thing you need to rely on in assessing the experimental result.

SHERRILYN ROUSH

Rice University