On the Testability of Psychological Generalizations (Psychological Testability)
Author(s): David K. Henderson
Source: *Philosophy of Science*, Vol. 58, No. 4 (Dec., 1991), pp. 586-606
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/188482
Accessed: 20/02/2009 18:46

ON THE TESTABILITY OF PSYCHOLOGICAL GENERALIZATIONS (PSYCHOLOGICAL TESTABILITY)*

DAVID K. HENDERSON†‡

Department of Philosophy
Memphis State University

Rosenberg argues that intentional generalizations in the human sciences cannot be law-like because they are not amenable to significant empirical refinement. This irrefinability is said to result from the principle that supposedly controls in intentional explanation also serving as the standard for successful interpretation. The only credible evidence bearing on such a principle would then need conform to it. I argue that psychological generalizations are refinable and can be nomic. I show how empirical refinement of psychological generalizations is possible by considering concrete cases. A sufficiently detailed view of the role of psychological generalizations in interpretation allows us to find in psychological investigations instances of bootstrap testing.

There is a long tradition in which challenges to the scientific status of social and psychological inquiry proceed by harping on the requirements of the interpretive understanding that would need to inform any such inquiry. Drawing on Davidson's thesis of the anomalousness of the mental, and, it seems, by reflecting on this venerable tradition, Alexander Rosenberg has produced a particularly challenging argument to the effect that psychological and social scientific generalizations cannot be law-like, at least to the extent that the human sciences use intentional idiom and seek to explain actions, beliefs, and attitudes.

The argument begins with the supposition that such inquiries would proceed on the basis of an intentional generalization linking belief and desire states as causes with actions as effects. Rosenberg envisions the following as representing the rudimentary form of such a generalization:

> [L] Given any person x, if x wants d and x believes that a is a means to attain d, under the circumstances, then x does a (Rosenberg 1988, 25; see also 1985, 402).

Of course [L] can be commonsensically refined. But the crucial question becomes whether or not [L] (or its commonsense refinements) can, in

*Received January 1990; revised May 1990.

†The author is indebted to Alexander Rosenberg, Terry Horgn, and John Tienson for helpful comments on earlier versions of this paper.

‡Send reprint requests to the author, Department of Philosophy, Memphis State University, Memphis, TN 38152.

Philosophy of Science, 58 (1991) pp. 586–606.

principle, be empirically tested and refined in a significant way. If such empirical refinement is impossible, then [L] (and its commonsense refinements) would be disanalogous to laws in the sciences generally; they would seem to be non-nomic (Rosenberg 1985, 36; 1988, 36, 47–49). For, "what is characteristic of science is its capacity to provide generalizations that are improvable beyond what common sense provides" (Rosenberg 1985, 404).

Rosenberg's argument can be stated as follows:

(1) There are only three broad ways that access to the intentional phenomena of interest in the human sciences and relevant to generalizations such as [L] might be obtained:
   (a) Typically, intentional states are attributed in an *[L]-dependent way*: interpretation. Interpretation is a matter of imposing a pattern on the beliefs and attitudes attributed to those interpreted. Significantly, the pattern is that indicated in [L]. Thus, in standard interpretation, attributed conformity with [L] becomes a measure of success.
   (b) One might imagine [L]-independent access to intentional states coming through neurophysical measuring instruments. This would require supporting psychophysical laws.
   (c) One might (perhaps) imagine behaviorist identification of intentional states.

(2) Behavioral approaches have proven to be obviously inadequate.

(3) There cannot be [L]-independent, interpretation-independent, measures of intentional states. The basis for this is Davidson's arguments for the anomalousness of the mental: There can be no psychophysical laws that serve to give us access to mental states of the sort at issue.

(4) LEMMA following from (1)–(3): *There really is only one way of getting at the intentional phenomena of interest: the [L]-dependent method of interpretive understanding. There are no [L]-independent ways.*

(5) If [L] is to be refined, there must be ways of comparing predictions or retrodictions derived using [L] with the relevant phenomena.

(6) In order to compare [L] with the phenomena of interest, we obviously must have ways of identifying the phenomena to see what obtains. But, comparability in the relevant sense requires [L]-independent access. For, on the (charitable) [L]-dependent procedure envisioned in (1a), failure to find conformity with [L] indicates problems with the interpretation, and thus, with the data that would supposedly serve as the basis for checking and refin-

ing [L]. Thus, there could be no phenomena (under interpretation) that indicated the need for refining [L].

(7) *Therefore,* on the basis of (4)–(6) [L] is not refinable.

As a result, Rosenberg concludes that psychological generalizations having to do with intentional states are not nomic causal generalizations, and that, "even though desire/belief explanations of actions do link causes to their effects [as Davidson has argued], their explanatory power does not consist in this fact" (1988, 40). Naturalists thus seem driven to espouse changing the traditional subject of the human sciences, becoming eliminativists. Antinaturalists, on the other hand, contentedly embrace a separatist understanding of hermeneutics.

Davidson's work, on which Rosenberg draws, makes much of the principle of charity in interpretation. Before evaluating the above argument, it is instructive to clarify the respects in which it does and does not depend on the principle of charity. Overall, and not considering the way in which particular premises rely on the principle of charity, the argument's general structure makes no use of that principle. It really turns on the claim that the principle that supposedly controls in intentional explanation, [L], also serves as the standard for successful interpretation. *Were we to suppose that some less charitable principle informed intentional explanation, this would not interfere with the general argument as long as we also suppose that the same principle also serves as the pervasive guide to the interpretation of the data on which its own appraisal rests.* To appreciate this point, consider a view of interpretation that is not fundamentally charitable. I have argued elsewhere that the fundamental constraint on interpretation is not charity, but rather a principle of explicability: interpret so as to find those interpreted to be explicable in their beliefs, attitudes, and actions (Henderson 1987a). Rosenberg's basic argument can be easily reconstructed employing the principle of explicability *if* we assume that some principle—[L] or some alternative—exercises as dominant an influence on intentional explanation as Rosenberg envisions for [L]. Supposing that [L] (or some alternative) constitutes the pervasive guide to intentional explanation, we would be led to take it as an equally pervasive guide to successful interpretation. The principle [L] (or the alternative) would then seem irrefinable, for no evidence could indicate a substantial need for refining it, as the only credible evidence would conform to it. At least, we can reason this way as long as we can suppose that there is no interpretation-independent way of identifying the relevant intentional states. Of course, this last claim is what premise (3) asserts.

It is in premise (3) that we find a reliance on Davidson's principle of charity in interpretation. For Davidson's arguments for the anomalousness of the mental rely on that principle. According to Davidson, there cannot be psychophysical laws because mental and physical predicates are "not made for one another" ([1970] 1980a, 218). When such predicates are combined in generalizations the results are said to be akin to that paradigm both of non-nomic generalizations and of combining predicates not made for one another: "All emeralds are grue". Davidson insists that mental and physical predicates are not made for one another because the central constraint on the use of mental predicates, namely, the principle of charity, has "no echo in physical theory" ([1974] 1980b, 231). Because I do not believe that the principle of charity is a fundamental constraint on interpretation, I am not convinced by this argument.

However, such reservations regarding the principle of charity and Rosenberg's third premise are not where I will concentrate my argument in this article. For, as a matter of fact, our stock of nomic psychophysical generalizations is none too impressive; at least it is not sufficient to yield anything resembling a "neurophysical measuring instrument" for beliefs and desires. Nevertheless, we do manage to make empirically motivated refinements in psychological generalizations. We do subject psychological generalizations to scientific tests. My goal here is to examine how we can manage this. My strategy is to concentrate on a supposition repeated several times in the preceding paragraph: There is a principle, [L] or some alternative, that exercises as dominant an influence on both intentional explanation and intentional interpretation as Rosenberg envisions for [L].

We can begin to appreciate our epistemic situation in the human sciences by considering what sorts of generalizations feature in psychological research in addition to Rosenberg's [L]. Obviously, [L] represents a rudimentary version of the ceteris paribus principles or generalizations that inform our rationalizing explanation of actions. As such, [L] can doubtless be refined somewhat to accommodate further aspects of our common sense rationalizing explanations. (Of course, insofar as refinements produce precise, unified accounts, these will typically incorporate elements that are not, strictly speaking, parts of our common sense.) Some would say that one prominent natural refinement is at least roughly that obtained by taking normative decision theory as informing rationalizing explanation. This much is suggested by Rosenberg (1988, 25–26, 65–74), and I will not here dispute the naturalness of formulations in decision theory, applied descriptively, as proposed refinements of [L]. I will refer to proposed refinements of [L] as [L^a].

It is important to notice the range of further generalizations that inform our explanations in the human sciences. Typically, proponents of rationalizing explanation also seek to explain intentional states themselves—people's beliefs, desires, and other attitudes. This involves the use of further principles (as generalizations) having to do with the relations between such intentional states. Let us denote this complementary set of generalizations [L^c]. However, not all intentional explanations are rationalizing explanations (Henderson 1989). Some are informed by general expectations for cognitive processing that is not rational, and some are informed by expectations for actions that are not rational in light of the agents' beliefs and desires. The relevant generalizations may be denoted by [G^a] and [G^c] respectively. Nothing in my reasoning will depend on categorizing the generalizations informing intentional explanations in this particular fourfold way. The purpose of the categories is simply to have some way to begin to reflect on the complexity of our resources outside of Rosenberg's [L] or even [L^a].

I believe that it is clear that we do manage to refine generalizations within each of the four categories, including [L^a]. We can, and do, contrive situations in which people's responses give us empirical reasons for revising our general expectations, including our [L^a]-expectations, thus refining [L]. Consider, for example, a set of situations and results produced by Amos Tversky (1975). The situations in question here are of the common sort involved in studies of decision making under uncertainty: choices between gambles. (Using the common notation, (X,P,Y) will represent a gamble where one will receive X with a probability of P, or Y with a probability of $1 - P$.) Subjects are presented with the choice between gambles A and B:

$$A = (\$1000, {}^1/_2, 0), \quad B = (\$400).$$

In situations of this first sort, almost all subjects prefer the "sure thing": B. They do this despite the fact that A has a greater expected actuarial value: \$500.

Of course, such results are not themselves news within standard decision theory. It is common to distinguish between the amount of goods or money to be had and its *utility*. The latter is conceived as a subjective, nonlinear, function of the former. The common postulation of decreasing marginal utility—a concave positive utility curve—is clearly enough to accommodate the results obtained in situations of this first sort: One need only claim that, for many subjects, $u(\$400) > {}^1/_2 u(\$1{,}000)$.

This standard response is just what Rosenberg would expect: Standard decision theory is not impugned by Tversky's results in the first situation because we interpret our subjects on the basis of its basic principles. We attribute to our subjects preference structures that conform to [L^a].

However, *we follow this line only to a point, after which such [L^a]-informed identification of values comes to clash with other constraints.* This begins to occur with the second situation, which is produced by multiplying the probabilities of the gains in A and B by ${}^1/_5$. Subjects are now asked to choose between:

$C = (\$1000, {}^1/_{10}, 0), \quad D = (\$400, {}^1/_5, 0).$

If the explanation of the choices found in the first situation were the concave shape of the subjects' preference curves, then we could expect these curves to produce preferences for D over C. However, most subjects now prefer C over D. As Tversky notes, within the confines of standard decision theory, the overall pattern of choices observed in the two situations is "incompatible with any utility function" (1975, 166).[1] Such results indicate a "*positive certainty effect* . . . [in which] the utility of a positive outcome appears greater when it is certain than when it is embedded in a gamble" (ibid.). Tversky goes on to provide evidence of a negative certainty effect when losses are in the offing instead of gains. Such interaction of utility and probability of options violates aspects of standard decision theory, where it is supposed that there are utility functions (unique up to a certain transformation) associated with particular goods and that such utility functions interact simply with subjective probabilities according to rule: $u(x)p(x)$. Thus, if Tversky's apparent finding of certainty effects is borne out, it will constitute a significant refinement within [L^a].

Tversky's results do seem to put a certain amount of empirical pressure on standard decision theory as [L^a]. To begin to appreciate how this is possible, we should consider several responses that might tempt proponents of descriptively applied decision theory.

Defenders of decision theory might suggest that Tversky's subjects just did not understand the situations in the way that he understood and presented the situations. In keeping with Rosenberg's argument, they might even cite Tversky's own results as evidence to this effect, arguing that investigators are obliged to interpret their subjects so as to find conformity with [L^a]. However, I think we are then entitled to a plausible account of just what it is about the situations that the subjects understood differently. Tversky's subjects were college students. Is it plausible to suppose that they do not understand talk of "dollars"? Or, that they did not understand the relative sizes of amounts such as 1,000 and 400? Is it then plausible to suppose that they did not realize that whatever common commodity can be purchased with \$400 can typically be gotten in a matched pair with \$200 left over from \$1,000? On these counts I believe that it is implausible to postulate misunderstanding. The reason, it seems, is that we have, in addition to [L^a], expectations regarding roughly when people learn rudimentary matters of importance within their society. Elementary arithmetic and monetary units are matters we expect are learned much

[1]Actually, one might suggest that the pattern obtained is the product of two utility functions, one having to do with monetary gains and losses, the other having to do with gambling and uncertain situations. However, this plausible suggestion is rendered problematic by the choices obtained in yet a third situation (Tversky 1975, 165–166).

before college age. *Such relatively mundane (but nevertheless empirical) expectations effectively block positing significantly different understandings of the relevant aspects of the situations presented.*

It may seem more plausible that the subjects understood the probabilities involved differently than did Tversky. But even this seems to me to be implausible. To begin with, the gambles were not made to depend on events regarding for which the subjects might have antecedently-acquired subjective probability estimates. Such antecedent probability estimates could conceivably serve as an anchor, affecting the subjects' subjective probability estimates even after the situation is stipulated by the investigator.[2] However, in the absence of such antecedent expectations, it is likely that their subjective probabilities conform to the stipulated probabilities, assuming that the subjects understand talk of a "probability of $^1/_2$", "of $^1/_5$", and "of $^1/_{10}$". Now, while many of the subjects may never have reflected on whether "probability" should be understood as a matter of relative frequency, or of propensity, or in both ways, and so on, it is not unlikely that they know well enough how to apply such descriptions to simple cases: coin tosses, urns with balls of mixed colors, the weather, their car starting on a cold morning, and so on. This would reflect enough shared understandings to make Tversky's results telling. (These relatively mundane expectations, like those above, are testable enough. So, the empirical case above could be buttressed somewhat.)

Again, we see that *in the particular situations produced here, we have expectations in addition to [L^a]-expectations that constrain us from gerrymandering our interpretive scheme too much.* What I have thought to appeal to here are socially applied crude generalizations regarding learning, something of a marriage of crude learning theory and information about socialization. Such presumably belongs to either [L^c], or to [G^c], or has components from both.

In a related vein, Tversky and Kahneman (1971) have produced a distinctly nonstandard decision theory, called "prospect theory", designed to accommodate a range of experimental results, including certainty effects. In prospect theory, there are "values", which are rather like utilities of outcomes (gains and losses) judged in terms of a neutral reference outcome. There are also "decision weights", which are nonlinear functions of probabilities. These are what is used to account for certainty effects. For, commonly, a weighting function, π, for an individual is such that $\pi(0) = 0$ and $\pi(1.0) = 1.0$, but a fair portion of the intermediate probabilities are undervalued: $\pi[P(x)] < P(x)$. Now, continuing the think-

[2]In keeping with the general position developed here, such suggestions for possible alternative construals of experiments are themselves based on empirical expectations. Anchoring effects are an empirically studied effect (Tversky and Kahneman 1974).

ing of the preceding paragraph, a defender of standard decision theory might propose taking these weighting functions as a way of representing empirically-determined subjective probabilities. After all, Tversky and Kahneman's choice problems typically stipulate probabilities and one might then take decision weights as representing what such and such a probability "feels like" to an individual. In view of the above, this would be to say that, although the subject may cognitively recognize simple applications of probabilities, decisions are made using a less well-behaved, almost gut-level, response to such probabilities. However, if one wished to retain most of the standard theory by calling the values achieved by way of these decision weights "subjective probabilities", then one would need to also hold that subjective probabilities need not be probabilities at all. (For, unlike probabilities, the decision weights attached to an event and its complement will often not sum to 1.0.) This would involve a rather substantial modification of [L^a] after all.

While Tversky's work discussed above may not lead inescapably to the conclusion that [L^a] must be modified, there is also a sense in which all empirical work is likely to be strictly inconclusive. What is more to the point is that this empirical work is generally compelling; it puts significant pressure on standard decision theory descriptively applied as [L^a]. The most straightforward interpretation of the subjects in the experiments leads us to attribute to them violations of principles of decision theory as [L^a]. We could abandon this interpretation of the experiment, but only at the cost of abandoning certain non-[L^a]-expectations as discussed above. As Quine has emphasized, we can hold any part of our web of beliefs constant—including [L^a]-parts—if we are willing to make radical enough compensating adjustment elsewhere in our beliefs. This sense in which any experiment is not absolutely decisive is the sense in which Tversky's is not decisive. However, often it seems unreasonable to make the adjustments elsewhere in our beliefs, and Tversky's work seems to put us in just this situation. In this situation our non-[L^a]-expectations—our mundane [L^c] and [G^c] resources—constrain us to retain the more straightforward interpretation.

What we find illustrated here is that a wider range of our expectations or generalizations constrain us in interpretation than is generally appreciated. This range is an epistemological resource of importance, for it allows us to test some generalizations using others in particular situations. We are then not left employing [L^a] alone, nor, for that matter, is [L^a] itself monolithic. Part of our set of general expectations can be played off against other parts in situations where those other expectations are relatively constraining. In this manner, [L^a]-generalizations, and others, can be empirically refined. This, in rough form, is the account that I propose to develop in the remainder of this paper. In effect, the central basis for the initial plausibility of Rosenberg's argument against the refinability of psychological generalizations is his ignoring the complexity or richness of our resources both in addition to [L^a] and internal to [L^a].

My central concern in this paper is to show that empirical results can be used to test the acceptability of particular intentional psychological generalizations, such as those in [L^a]. To this end, I now draw on Clark Glymour's bootstrapping account of scientific testing, a promising account of how empirical results can be evidentially relevant to particular parts of our theories. The ability to distribute praise and blame has proven difficult to account for in terms of the Hypothetico-Deductive approach. Glymour takes accounting for it to be an adequacy condition on accounts of scientific evidence and testing (1980, 3). He shows that his own "bootstrapping" account succeeds here, and argues that it provides a basis for an accurate understanding of some (but not all) important episodes in the history of science (ibid., 169–172).

According to the bootstrapping account of scientific evidence and testing, data provides evidence for theoretical claims, hypotheses, by providing instances of the hypotheses (where an "instance" is understood in something like the way developed in Hempel's work on qualitative confirmation). Of course, when the hypothesis under test employs vocabulary not used in statements of the data, a very common situation, the data can provide an instance of the hypothesis only when an instance of the hypothesis can be derived from the evidence in conjunction with further theoretical claims. Thus, confirmation becomes a triadic relation: Certain data confirms (or disconfirms) hypothesis H with respect to theory T (Glymour 1980, 110). The theoretical claims connecting data to hypothesis are not here thought to be analytic truths. Rather, they are themselves subject to testing (confirmation or disconfirmation) against data with respect to some bits of theory (ibid., 145–151) (thus the notion of bootstrapping).

Glymour's account raises a familiar worry—mutually compensating errors:

> Suppose that bit A is used together with evidence E to justify bit B; and suppose that bit B is used with evidence E' to justify bit A; then might it not be the case that E and E' determine neither a real instance of A nor a real instance of B but provide instead *spurious* instances, which appear to be correct because the errors in A are compensated for in these cases by errors in B and vice versa? (Ibid., 108)

Glymour responds, "Indeed it might be the case, but it would be wrong to infer straight out from this that E and E' provide no grounds at all for A and B" (ibid.). He seeks to build into his account ways of guarding against compensating error and concludes that "[t]he only means available for guarding against such errors is to have a variety of evidence, so that as many hypotheses as possible are tested in as many different ways as possible" (ibid., 140). Such variety in testing any one hypothesis involves computing the values involved in that hypothesis in different ways, using different sets of theoretical generalizations and computing them from different sets of evidence.

To test a hypothesis, H, on the basis of evidence E with respect to theory T, then, it must be possible to compute or determine the value of the quantities occurring in the hypothesis so as to obtain a positive instance of it on the basis of E and generalizations of T.[3] However, while this is a necessary condition, it is clearly not sufficient for testability for it is also necessary that evidence of the sort E puts H at risk. That is, E contains values for certain quantities (understood broadly), and it must be possible to have gotten from the procedures used in obtaining E a data set E' with values of those same quantities that would, when used in the same computations that produced the positive instances, produce a negative instance of H (Glymour 1980, 115–117).

Rosenberg's worry is just that all possible attempts at finding negative or positive instances of [L^a] fail this last requirement of bootstrap testing. In particular, his argument can be understood as an attempt to show that the following undesirable state of affairs, as described by Glymour, characterizes [L] and [L^a], "it may turn out that some hypothesis or other has to be used in every computation in such a way that that hypothesis itself cannot be tested from the evidence" (ibid., 134). This ultimately does not characterize the epistemic situation of [L^a], but the case has to be made with care. We will see that a bootstrapping model can incorporate the observations already made in connection with Tversky's illustrative work.

First, however, we need to recognize one respect in which the basic characterization of bootstrap testing provided so far is clearly too demanding. I have followed Glymour's original presentation in requiring that the values of the quantities in the hypothesis be completely determinable from the evidence together with the relevant additional theoretical generalizations. Now, it is clear that this requirement is very seldom, if ever, met in the human sciences. However, the human sciences are certainly not alone here. This has prompted van Fraassen (1983) to provide a flexible reconstruction of Glymour's account in which it is not supposed that the values of the quantities in the hypothesis must be completely or exactly determinable in order for that data to provide even a moderate test of the hypothesis with respect to the relevant theory. It is

[3]This statement can be understood or modified to apply rather naturally to hypotheses that are not equations (Glymour 1980, 123–133).

necessary to allow for approximations in which the ranges of values for the quantities in the hypothesis are restricted sufficiently. Here a range of somewhat weaker positive tests are achieved when the ranges obtained are fully compatible with the hypothesis, while different possible evidence would have produced ranges incompatible to the hypothesis. Such testing is clearly in evidence in the work by Tversky already discussed, where we find appeals to what is "plausible" and "implausible" to attribute to subjects, and talk of what is and what is not "likely". It is worth noting that such a narrowing of ranges of values for quantities of interest is also found in the interpretation of experiments in physics, for example. This is clearly seen in accounts of experiments devoted to the detection of solar neutrinos (Pinch 1986, Shapere 1982).

Of course, provision must also be made for developing theories in which "more and more quantities may become empirically determinable; conversely, a good deal of testing is already being done, while some of the crucial quantities still remain indeterminate" (van Fraassen 1983, 32). Indeed, Glymour's own discussion of atomic theory in the nineteenth century reflects ways in which progress can be made in such situations (Glymour 1980, 226–263).

Alison Wylie has convincingly argued that certain research in archaeology "does conform to Glymour's model in intent and, in broad outline at least, in practice" (1986, 314–315). Such a guarded positive conclusion can also be drawn for important work in cognitive psychology. However, the implementation of bootstrap testing here is complicated somewhat by the much remarked-on holistic character of intentional states.

To appreciate the complications, consider the failure of a naively straightforward application of the bootstrapping model. We may begin by asking what sort of information would count as data and what sorts of predicates would count as representing quantities whose values are to be determined. The most plausible candidates for data are characterizations of behavior, verbal and otherwise. The hypotheses to be tested are supposedly generalizations about intentional states, and intentional states and actions. Providing instances of such generalizations will then be a matter of attributing states (or ranges of states) to subjects on the basis of their utterances and other behavior. To do this, we supposedly would need further generalizations that connect the observed behaviors with the intentional states mentioned in a given hypothesis. This naive bootstrapping model of psychological testing would have us proceed as follows: We observe some behavior in the course of a particular experiment. Then, we employ a set of relevant psychological generalizations to that rather limited set of observations in order to derive intentional attributions that cumulatively constitute either negative or positive instances of some hypothesis stated in intentional terms.

The problem with this naive approach is that there are no plausible candidates for the role of interesting psychological generalizations that will get us from a set of characterizations of behavioral events to characterizations of intentional events and states. At least this is so when we look for generalizations that would connect relatively discrete sets of behavior with intentional states—for example, grabbing behavior with wanting to have the thing grasped. Does the battlefield hero really want the grenade that he grabs up from among his fellows? Such an approach to applying the bootstrapping model is entirely too behavioristic (not in the sense of a methodological program that would forego intentional idiom, but in the sense of a program in the tradition of philosophical behaviorism that would take relatively discrete behaviors as indicative of mental states).

As already indicated, the reason for the failure of the naive approach is the holistic character of the mental. But, such holism need not foreclose producing instances of intentional generalizations on the basis of psychological generalizations and data that is more or less behavioral. It is just that, in keeping with it, psychological generalizations will have to be put to work in a much more cooperative manner than that envisioned above. My model here is that of theory-laden translation and interpretation (Henderson 1987a). At the most general level, the relations between the holism of the mental, interpretation, and our psychological generalizations can be understood as follows. The holism of the mental is basically a matter of the type and content of intentional states depending on their place in a pattern. Thus, whether we can attribute a particular state to a person may well depend on what other states we attribute to that individual, as well as on the person's behavior. The cooperative use of psychological generalization is, in effect, the basis for the pattern we seek to find in interpretation (Henderson 1989). In this way, a range of our psychological generalizations provide a basis for interpretation, and thus for the provision of instances from more or less behavioral data. If I am correct, we test and refine psychological generalizations by producing negative or positive instances of such generalizations under interpretation when interpretation is both holistic and theoretically informed.

The challenge in developing this general position is obvious: An adequate model of bootstrapping in psychology must reveal how the holistic nature of interpretation *does not* result in "some hypothesis or other", say those of [L^a], needing "to be used in every computation in such a way that that hypothesis itself cannot be tested from the evidence" (Glymour 1980, 134). It must provide for holistic, theory-laden interpretation that does not result in the theory informing the interpretation being self-verifying in the way Rosenberg envisions.

In developing such a model, we should consider two intimately related tasks: the construction of translation manuals or interpretive schemes generally and the application of such schemes in empirical work. A proper account of the first matter will allow us to subsume the second matter as a special case in a way that provides for bootstrap testing in psychology.

When considering the construction of interpretive schemes, it is useful to focus initially on what Quine and Davidson would call "radical translation" (or "radical interpretation"). By supposing that we build our interpretive scheme from scratch, without the help of intermediaries, we lay bare the more fundamental constraints on interpretation. I believe that what we find is that we should seek to attribute to our subjects intentional states, beliefs and desires, and actions that are patterned in a certain general way and that thereby account for subjects' behavior as a whole (at least so far as we can sample it) in light of our psychological expectations. Thus, significantly, the relevant pattern is dictated by our general psychological expectations. (The role of [L^a]-informed expectations here will need to be considered with particular care.) At this general level of analysis, interpretation is naturally viewed as the theory-based attribution of states, and, in this respect, fits nicely into the bootstrapping account, as suggested above.

In most cases, much of our intepretation will proceed in terms of the application of a translation manual or interpretive scheme. (In Tversky's work that is discussed above, the implicit translation manual was basically homophonic.) In these cases, we are employing an empirical scheme which has been produced in the course of interpretive work in the relevant culture. This work has found that, in the society studied, certain behaviors are generally indicative or expressive of certain intentional states under interpretation. That is, such work can be understood as finding certain utterances expressive of states that, when attributed to utterers in a society, fit their behavior (at least so far as we can sample it) into a general pattern informed by our psychological expectations. Such a manual does not stand in place of interpretation; rather, it gives us a running start at interpretation. I will consider these two sorts of cases in turn.

If we are to overcome Rosenberg's worries, it is important to distinguish between the earlier and later stages of constructing a translation manual or interpretive scheme. I have argued elsewhere that such a distinction is required by anthropological practice and reflects significant differences in the constraints facing actual investigators (Henderson 1987a). In the earlier stages, we seek to construct a first-approximation scheme. In this task, the application of sophisticated theory informing interpretation is pretty much out of the question. Rather, we seek to attribute beliefs and desires (and actions) in a way that accords with rudimentary expectations, and expectations whose antecedent conditions require little information to determine whether they hold. Interpretive practice in the early stages turns out to be "charitable" because we clearly expect correct judgements in simple perceptual situations dealing with objects familiar to our informants and we expect rudimentary rationality both in the form of elementary logic applied to some simple cases and in the form of practical rationality in limited application. I thus understand charity in interpretation to be empirically founded, as does Quine (1987, 7).

Having generally satisfied the above constraints in the early stages of interpretation, we have access to presumptively or generally correct information about a fair range of our informants' beliefs and desires. With this in hand, we can go on to apply more and more sophisticated psychological theory to address more adequately the explicability of surprising beliefs, desires, or actions that we occasionally find ourselves led to attribute to those who are being studied. This is the essence of the later stages in translation or interpretation. If we find that our subjects are explicable as interpreted according to our first-approximation scheme, this reinforces the scheme and adds another, if somewhat yet tentative, confirming instance for the relevant theory. If, however, certain phenomena under first-approximation interpretation prove recalcitrant to explanation using more sophisticated theory, we face the choice: Modify the interpretive scheme at some relevant point (taking care not to create other recalcitrant anomalies), treat the case as a negative or disconfirming instance of the relevant theory, or employ a mixed response—placing less confidence in both scheme and theory.[4]

For our purposes here, the importance of the distinction between the earlier and later stages in constructing interpretive schemes is that, without employing sophisticated theoretical expectations, we can in the reasonably early stages of interpretation learn a good amount about a people's categories, beliefs, and values. Although some cases will remain puzzles awaiting further investigation, we can generally learn a good deal about what sorts of things they take to be consequences of what other sorts of things, and about what they take to be the case. Of particular importance for our concerns here, we can learn a fair amount about *how they express preferences and choices,* and about some of their preferences.

Again, it must be emphasized that this can be done without using sophisticated theory, such as variants of [L^a]. The more rudimentary patterns reflected in common everyday explanations seem to suffice in many cases. In association with identifying source-language expressions for wants and desires, we typically seek to understand indigenous explanations of behaviors, source-language users' own requests, and expressions of choices.

[4]These remarks are not meant to suggest that the rudimentary theory used in the earlier stages of interpretation are themselves immune to revision. Revision there is admittedly less likely. But, if efforts at less drastic measures in the later stages are repeatedly unsatisfactory, attention could be focused on the more fundamental expectations used earlier.

(Of course, we need to be able to translate our informants' categorizations of what is wanted.) In this interpretive endeavor, we seldom need to construct elaborate representations of our subjects' utility preferences, "unique up to a linear transformation", as they say in decision-theoretic work. (Indeed, I can think of no ethnography that is supported by a demonstration that the interpretive scheme used leads us to attribute beliefs and utilities in anything resembling full keeping with decision-theoretic expectations.) Instead, we seek to construct a rather rudimentary explanatory model of those who we interpret, one that allows us to attribute some enduring or "standing" wants (wanting to acquire horses, wanting a spot in close to the river to camp, wanting to be admired) and "occurrent" wants (wanting a drink, wanting that horse), and all this in a way that allows us to make use of rudimentary explanations. Often, but not exclusively, these simple explanations will be rudimentary rationalizing explanations. While this will sometimes involve some information on the relative strength of the wants we attribute to those we study, typically it will not involve much precision here. Much might be (and typically is) done without anything but crude ranges into which we have sorted the desires of our subjects. If we can do this much, it is *enough to give us the confidence that our interpretive scheme generally allows us to determine what is wanted* (assuming a modicum of candor, of course) *and to determine when a choice is made*. The crucial point is that these accomplishments seem to be attainable without the use of anything of a sophisticated nature in the way of [L^a].

When we investigate the thoughts and thought processes of people, or of a particular group of people, by employing an interpretive scheme for their society, we are engaging in the later stages of translation or interpretation. Tversky's study discussed above can be understood in this way. In practice, Tversky relied on an implicit first-approximation translation manual or interpretive scheme. I think it is safe to say that his implicit scheme measures up to the standards for a serviceable first-approximation scheme. (And, doubtless, he designed his experiment so that the threat of linguistic divergence was minimized by a selective use of the supposed translation scheme.) Thus, there is a prima facie case for the general reliability of his characterizations of his subjects' choices. And, as we have seen, further considerations support this presumptive case. This is typical of the later stages of interpretation as described above, for there further theory is typically brought to bear in determining the plausibility of possible modifications (or nonmodification) of the first-approximation scheme employed.

The above way of conceptualizing matters has important implications for the possibility of bootstrap testing of [L^a]-generalizations. For it turns out that [L^a] must now be understood as internally complex, having a rudimentary core that does not go much beyond [L] together with rough provision for competing desires. Such a core is then elaborated within [L^a] by clauses of various sorts, stipulating the shape and nature of utility curves perhaps, or the nature of subjective probabilities, because [L^a] is internally complex, something less that the whole of it can be used in arriving at a basic, or first-approximation, interpretive scheme. Only a rudimentary component of it needs to be used or supposed in the interpretive schemes on which tests of sophisticated variants of [L^a] are based. Thus [L^a] *is not so fully involved in the computation of values—the attributions of intentional states—that it interferes with its own testing*.

At the very least, Tversky's accomplishment is obtaining a set of choices (under interpretation) that are incompatible with the principles of standard decision theory as discussed earlier. He can "calculate" roughly the conditions of interest—choices, preference—cautiously using his homophonic manual, buttressed by further empirical considerations (generally taken from [L^c] and [G^c]). It then seems infeasible to find in the choices actually obtained an instance of the hypothesis under test: standard decision theory applied as [L^a]. Thus, a negative test.

One possible response is to suggest that, while refinements of [L] may be testable in the way I have suggested, [L] itself is not testable. But such a response simply misses the point. It would admit that [L] *is empirically refinable* in the sense that elaborations on it that are empirically stronger and more precise could be tested. If any elaboration were to prove empirically adequate (or even a substantial improvement) we would certainly want to employ that principle in scientific contexts. Thus, the claim that [L] itself is an untestable core principle of intentional psychology would not preclude there being scientific refinements in psychology. (Rosenberg 1988, 47 would agree that refinability is the central issue.)

However, the claim that [L] is untestable is mistaken. It is reminiscent of Davidson's ([1970] 1980a, [1974] 1980b) claim that it is a priori that beliefs and desires are preponderantly rational. Davidson makes this a conceptual matter: If we do not attribute preponderant rationality to those we interpret, we would not be using talk of beliefs and desires in the standard way, and we would thereby "change the subject". Of course, this would not ensure that people are preponderantly rational. It shows at most that what beliefs and desires people have (and they may yet have none) are preponderantly rational. Similarly, if the claim that [L] is untestable is the claim that we could never be led to abandon using [L] in psychology, then the argument is not sound, and its conclusion is probably false. However, if we were to abandon [L], it is plausible that we would no longer be doing intentional psychology, *strictly* speaking. From the perspective of much of psychology and the human sciences, this would be to change the subject. But, as long as we are being strict about what counts as the subject of intentional psychology, we should also recognize that the new subject that results need not be all that different from intentional psychology as presently understood. We might have as internal states "B-states" and "D-states" that are recognizable as conceptual descendants of "beliefs" and "desires".

If my argument is correct to this point, we can conclude first that [L] is refinable, for refinements of [L] are testable, and second that [L] itself is testable in the sense that we could be led to finally give up on intentional psychology insofar as it is understood in terms of [L].

My discussion to this point has focused on one way in which the complexity of the generalizations underlying the rationalizing explanation of actions provides a basis for tests and refinements of such generalizations: Only a rudimentary core is involved in the provision of a first-approximation interpretation or translation. Relying on this, one can produce enough of an interpretation to justify the conclusion that no positive instance of the candidate refinement can be obtained in some concrete case. Thus, a proper part of [L^a] can be used in testing putative refinements on it.

Other cases of parts of [L^a] being used to test distinct parts are certainly conceivable. Suppose that certain refinements on [L] have proven empirically supported and have come to be accepted as reasonably well confirmed. We might then use these as a constraint on further interpretation. Thus we might encounter a case in which we face a choice: attribute a violation of this well-confirmed refinement on our expectation, or produce an interpretation that is a disconfirming instance of another proposed refinement. In this case, it is reasonable to use the well-confirmed refinement as a part of the theory with respect to which the proposed additional refinement is tested. We then "calculate" a negative instance of the hypothesis, the proposed additional refinement. These reflections suggest that refinements in psychological generalizations add to the resources we have in psychological testing. This is simply an instance of what van Fraassen described as a "developing theory be[coming] ever *more testable*" (1983, 33).

Having now sketched the way in which the internal complexity of [L^a] allows us to use parts of it to test other parts of it, we should explicitly note the place for psychological generalizations other than those in [L^a] in providing for psychological testability—that is, the place for generalizations from [L^c], [G^a], and [G^c]. All such generalizations can contribute to the testability of other psychological generalizations in the capacity assigned above to [L] and to well-confirmed [L^a]-generalizations: They can provide constraints on interpretation. (This role was already in evidence in the concrete psychological studies discussed earlier.) In this way, they provide resources for the "determination" of states mentioned in yet other proposed psychological generalizations, thus becoming part of the theory against which hypotheses can be tested. For example, refinements in our general expectations concerning human inductive reasoning can serve to make psychological theory more testable by serving in turn as a constraint on interpretation.

Further, such refinements may be obtained in much the way that we have found for refining [L^a]-expectations. Thus, if, for example, one were to consider recent psychological work on the sort of heuristics posited by Kahneman and Tversky (1972) and Tversky and Kahneman (1974) one would find that those investigations (and their apparent positive tests) take a now familiar form: An implicit first-approximation interpretive scheme is employed here, and a rich enough set of data is elicited to allow checking for some possible deficiencies. The questions addressed to the subjects are designed so that if the subjects' understandings of the tasks given them differ from the investigators' understanding (based on the implicit translation manual), then some clue or hint to this effect should appear in the pattern of responses. Additionally, one could point to the relatively straightforward nature of the task given to the subjects and ask just what it was that was supposedly not understood. Again, we would find a set of relatively mundane (but empirical) expectations strengthen our confidence in the implicit first-approximation interpretive scheme employed.

Thus, the account developed focusing on [L^a]-generalizations can be applied *mutatis mutandis* to the testability of these other classes of generalizations and to their role in the testability of yet other generalizations.

Before summarizing my results, I must acknowledge what seems to me the most serious limitation of the model of psychological testing developed here. It will be remembered that the one limited defense against mutually compensating errors in bootstrap testing was to "have a variety of evidence, so that as many hypotheses as possible are tested in as many different ways as possible". But the holism of the mental and what I called the cooperative application of psychological generalizations lead to limitations on the variety that is attainable here. One way in which tests are relevantly different is in using "different calculations", resting on different sets of supporting theory, in producing instances of hypotheses. However, attaining a basic first-approximation interpretive scheme seems to rely on the constraints imposed by a characteristic set of rudimentary theory that is applicable in the early stages.[5] Thus, all tests will apparently be similar in resting, in part, on this rudimentary core. However, even so, some place for variety of tests remains, for significant

[5]I do believe that this theory can vary over time. But at any one time the set we would draw on might be reasonably constant.

differences will arise in the sort of tests described here as, in different cases, different additional generalizations become particularly relevant as further constraints on interpretation.

In general then, we find that an appreciation for the complexity of our set of psychological generalizations, combined with an adequate account of interpretation as a theory-informed endeavor, allows us to account for the testability of psychological generalizations as a matter of bootstrap testing. In particular we find that empirical refinements in [L^a]-generalizations are possible when we take into account how a generally adequate first-approximation interpretation can be developed that does not rely on those refinements in attributing the relevant states to subjects. Further, we find that a range of other types of psychological generalizations can similarly be tested and can be resources in psychological testing. Theory-informed interpretation here amounts to the determination of values of the "quantities" mentioned in the psychological hypothesis being tested. A negative test occurs when a rich set of (roughly behavioral) data is produced that (a) when straightforwardly interpreted according to a first-approximation scheme, produces a negative instance of the hypothesis, and (b) cannot be reinterpreted to provide a positive instance without violating other psychological generalizations.

In fairness to Rosenberg, I should mention an aspect of his interesting argument that has been neglected to this point. I have considered his argument (as formulated at the start of this article) on its own internal merits, challenging the premised understanding of interpretation, and arguing that a close examination of our epistemic resources in psychological work reveals leverage for scientifically refining [L]. However, Rosenberg (1988, 30) seems to conceive of his argument, premises and all, as at least in part supported by what it allows us to explain: the alleged lack of progress in the social sciences when compared with progress in the natural sciences. This relative lack of progress is to be explained as the result of the social sciences relying on folk psychology through much of their history, while folk psychology is said to be wedded to empirically irrefinable commonsense variations on [L]. Rosenberg's argument, including its premises, is thus to be supported by an inference to the best explanation. Rosenberg can then justly demand of me whether I would deny that the explanandum phenomenon obtains, and, if I do not, what alternative explanation I could supply.

I believe that the appropriate response is a guarded one along three lines. First, inferences to the best explanation are appropriate only when the potential explanation that is to be supported in this way is not just the best among present competitors, but is also a meritorious potential explanation itself. But, if a potential explanation supposes that a certain state of affairs obtains when there is good reason to believe that it does not, then that potential explanation as it stands can hardly be meritorious (although some related potential explanation might yet prove meritorious). Thus, the fact that there is reason to believe that a potential explanation relies on false suppositions renders that explanation not meritorious and not supportable by argument to the best explanation, even if it is the best explanation we presently have. Now I have here presented reasons for believing that the explanation Rosenberg advances for the paucity of social scientific progress rests on a faulty view of [L] as not empirically refinable. I have described our epistemic situation in a way that allows us to appreciate how such principles can (in principle) be empirically refined. Further, I have discussed cases where empirical pressures are brought to bear on such principles. Thus it seems that Rosenberg's argument does not provide us with a meritorious explanation.

Second, to the extent that the social science and even psychology have seen little progress, compared to the natural sciences, this seems to me best explained by noting a range of complementary factors. To begin with, rather than claiming that the central generalizations of intensional psychology are wholly irrefinable by empirical work, one might account for the *relative* lack of progress by noting that the generalizations are *relatively difficult* to so refine. In view of my discussion above, this would seem to be a reasonable conclusion. In addition, other factors surely contribute to the frustratingly slow pace of social scientific progress. Obviously, certain social scientific debates have been initiated and perpetuated by concerns to protect or further certain ideological sacred cows; this has doubtless hindered progress. Further, some social scientists have taken a studied conservative stance to the variants of intensional psychology that they might draw on as psychological work does progress; they thereby have treated common rationalizing explanation and [L] as privileged, if not sacrosanct. Other factors, of course, might also be cited here. Together, these observations seem to me to comprise the best present explanation of the paucity of social scientific progress.

Third, while progress may have come relatively slowly to psychology, and to the social sciences in particular, it seems to me to have come. And empirical work has proved an inducement. Cultural anthropology *has* progressed from the days of Frazer and Taylor. And empirical work (properly understood) has played its part. With extended field work, observations showed that earlier accounts of magical and religious belief as socially elaborated childishness were inadequate. Observations also showed simple symbolist accounts inadequate (Geertz 1973, 101; Henderson 1987b). More directly to the point, my concern here has been primarily with the pivotal case of psychology, and it seems to me that empirical work in cognitive psychology, such as that discussed earlier, is resulting in empirical pressure being exerted on traditional psychological principles. The results are proposed refinements. Some of these may well survive further empirical work, and this, so far as I can see, is what empirical refining generalizations largely amounts to.

REFERENCES

Davidson, D. ([1970] 1980a), "Mental Events", in D. Davidson, *Essays on Actions and Events*. Oxford: Clarendon Press, pp. 207–225.

——— ([1974] 1980b), "Psychology as Philosophy", in D. Davidson, *Essays on Actions and Events*. Oxford: Clarendon Press, pp. 229–244.

Føllesdal, D. (1986), "The Status of Rationality Assumptions in Interpretation and in the Explanation of Actions", in E. LePore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell, pp. 311–323.

Geertz, C. (1973), "Religion as a Cultural System", in C. Geertz, *The Interpretation of Cultures; Selected Essays*. New York: Basic Books, pp. 87–125.

Glymour, C. (1980), *Theory and Evidence*. Princeton: Princeton University Press.

Henderson, D. (1987a), "The Principle of Charity and the Problem of Irrationality (Translation and the Problem of Irrationality)", *Synthese 73*: 225–252.

———. (1987b), "Winch and the Constraints on Interpretation: Versions of the Principle of Charity", *Southern Journal of Philosophy 25:* 153–173.

———. (1988), "Wittgenstein's Descriptivist Approach to Understanding: Is There a Place for Explanation in Interpretive Accounts?" *Dialectica 42*: 105–115.

———. (1989), "The Role and Limitations of Rationalizing Explanation in the Social Sciences", *Canadian Journal of Philosophy 19*: 267–287.

Kahneman, D. and Tversky, A. (1972), "Subjective Probability: A Judgment of Representativeness", *Cognitive Psychology 3*: 430–454.

Nisbett, R. and Ross, L. (1980), *Human Inference: Strategies and Shortcomings in Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Quine, W. (1987), "Indeterminacy of Translation Again", *Journal of Philosophy 84*: 5–10.

Pinch, T. (1986), *Confronting Nature: The Sociology of Solar-neutrino Detection*. Dordrecht: Reidel.

Rosenberg, A. (1985), "Davidson's Unintended Attack on Psychology", in E. LePore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell, pp. 399–407.

———. (1988), *Philosophy of Social Science*. Boulder: Westview.

Shapere, D. (1982), "The Concept of Observation in Science and Philosophy", *Philosophy of Science 49*: 485–525.

Tversky, A. (1975), "A Critique of Expected Utility Theory: Descriptive and Normative Considerations", *Erkenntnis 9*: 163–173.

Tversky, A. and Kahneman, D. (1971), "Belief in the Law of Small Numbers", *Psychological Bulletin 76*: 105–110.

———. (1974), "Judgment under Uncertainty: Heuristics and Biases", *Science 185*: 1124–1131.

van Fraassen, B. (1983), "Theory Comparison and Relevant Evidence", in J. Earman (ed.), *Minnesota Studies in the Philosophy of Science*. Vol. 10, *Testing Scientific Theories*. Minneapolis: University of Minnesota Press, pp. 27–42.

Wylie, A. (1986), "Bootstrapping in the Un-Natural Sciences: Archaeological Theory Testing", in A. Fine and P. Machamer (eds.), *PSA 1986*, vol. 1, pp. 314–321.