Rationality and Theory Choice

"Rationality and Theory Choice" was presented to the American Philosophical Association at a symposium on the philosophy of Carl G. Hempel in December 1983. The proceedings of the symposium were published in The Journal of Philosophy *80 (1983). Reprinted with the permission of* The Journal of Philosophy.

THE REMARKS THAT FOLLOW are a much compressed status report on one product of my continuing interaction with C. G. Hempel. That interaction began twenty years ago with my arrival at his University and my middle years. If new masters can be acquired at that age, then Hempel became mine. From him I learned to recognize philosophical distinctions centrally relevant to my enterprise. In him I learned to recognize the stance of a man who intends philosophical distinctions to advance truth rather than to win debates. Participation in a symposium that honors him gives me much pleasure.

Among the topics that have prompted frequent and lively exchanges between us is the evaluation of and choice between scientific theories. More than other philosophers of his persuasion, Hempel has examined my views in this area with care and sympathy: he is not one of those who suppose that I proclaim the irrationality of theory choice. But he sees why others have supposed so. Both in writing and in conversation, he has underscored the lack of argument or of apparent concern with which I switch from descriptive to normative generalizations, and he has repeatedly wondered whether I quite see the difference between

The final revisions of this paper owe much to the critical intervention of Ned Block.

explaining behavior, on the one hand, and justifying it, on the other.[1] It is to our continuing discussion of these questions that I now return. Under what circumstances may one properly claim that certain criteria which scientists are *observed* to use when evaluating theories are, in fact, also rational bases for their judgments?

I begin with a suggestion that I originally developed when commenting on a paper of Hempel's at Chapel Hill in 1976. Both he and I premise that the evaluation of criteria for theory choice requires the prior specification of the goals to be achieved by that choice. Now suppose—a simplistic assumption which will later prove dispensable—that the scientist's aim in selecting theories is to maximize efficiency in what I have elsewhere called "puzzle solving." Theories are, on this view, to be evaluated in terms of such considerations as their effectiveness in matching predictions with the results of experiment and observation. Both the number of matches and the closeness of fit then count in favor of any theory under scrutiny.

Clearly, a scientist who subscribed to this goal would be behaving irrationally if he sincerely said, "Replacing traditional theory X with new theory Y reduces the accuracy of puzzle solutions but has no effect with respect to the other criteria by which I judge theories; nevertheless, I shall select theory Y, setting X aside." Given the goal and the evaluation, that choice is obviously self-defeating. Similar considerations apply to a choice of theory whose *sole* effect with respect to criterial measures is to reduce the number of puzzle solutions, to decrease their simplicity (thus making them harder to achieve), or to increase the number of distinct theories (and thus the complexity of apparatus) required to maintain the puzzle-solving capacities of a scientific field. Each of these choices would be in prima facie conflict with the professed goal of the scientist who made it. There is no clearer sign of irrationality. Arguments of the same sort can be developed for other standard desiderata invoked when evaluating theories. If science can justifiably be described as a puzzle-solving enterprise, such arguments suffice to prove the rationality of the observed norms.

Since our encounter at Chapel Hill, Hempel has on occasion suggested what I take to be a deeper version of the same point. In the penultimate paragraph of a paper published in 1981, he points out that some of the difficulties with my published accounts of theory choice

1. See, for example, his "Scientific Rationality: Analytic vs. Pragmatic Perspectives," in *Rationality Today*, ed. Theodore F. Geraets (Ottawa: University of Ottawa Press, 1979), pp. 46–58.

would be avoided if desiderata like accuracy and scope, invoked when evaluating theories, were viewed not as means to an independently specified end, like puzzle solving, but as themselves goals at which scientific inquiry aims.[2] More recently still, he has written:

> Science is widely conceived as seeking to formulate an increasingly comprehensive, systematically organized, world view that is explanatory and predictive. It seems to me that the desiderata [which determine the goodness of a theory] may best be viewed as attempts to articulate this conception somewhat more fully and explicitly. And if the goals of pure scientific research are indicated by the desiderata, then it is obviously rational, in choosing between two competing theories, to opt for the one which satisfies the desiderata better. . . . [These considerations] might be viewed as *justifying* in a near-trivial way the choosing of theories in accordance with whatever constraints are imposed by the desiderata.[3]

Because it loosens the commitment to any particular prespecified goal like puzzle solving, Hempel's formulation is an improvement on mine; our points are otherwise the same. But if I read him correctly, Hempel is less satisfied than I with this approach to the problem of the rationality of theory choice. He refers to it as "near-trivial" in the passage just quoted, apparently because it rests on something very like tautology, and he finds it correspondingly lacking in the philosophical bite one expects from a satisfactory justification of the norms for rational theory choice. In particular, he underscores two respects in which near-trivial justification seems to fail. "The problem of formulating norms for the critical appraisal of theories may," he points out, "be regarded as a modern outgrowth of the classical problem of induction," a problem that the near-trivial justification "does not address at all" (92). Elsewhere he emphasizes that, if norms are to be derived from a description of the essential aspects of science (my "puzzle-solving enterprise" or his "increasingly comprehensive, systematically organized, world view"), then the choice of the description that serves as premise for the near-trivial

2. "Turns in the Evolution of the Problem of Induction," *Synthese* 46 (1981): 389–404. This position is foreshadowed on p. 42 of the paper cited above, where Hempel notes the difficulties in deciding whether a particular desideratum, e.g., simplicity, should be viewed as a goal or as a means to its attainment.

3. "Valuation and Objectivity in Science," in *Physics, Philosophy and Psychoanalysis: Essays in Honor of Adolf Grünbaum*, ed. R. S. Cohen and L. Laudan (Boston: Reidel, 1983), pp. 73–100; quotation from pp. 91 f. Subsequent references to this paper will be by parenthetical page number in the text.

approach itself requires justification which neither of us appears to provide (86 f., 93). The activities observed by a science watcher can be described in countless different ways, each the source of different desiderata. What justifies the choice of one of these, the rejection of another?

These examples of the shortcomings of the near-trivial approach are well chosen, and I shall shortly return to them. At that point I shall sketch an argument suggesting that a particular sort of descriptive premise requires no further justification and that the near-trivial approach itself is therefore deeper and more fundamental than Hempel supposes. In doing so, however, I shall be venturing into what is for me new territory, and I want first to clarify the argument by indicating its relation to positions that, for another territory, I have developed in some detail before. If I am right, the descriptive premise of the near-trivial approach exhibits, within the language used to describe human actions, two closely related characteristics that I have previously insisted are essential features also of the language used to describe natural phenomena.[4] Before returning to the problem of rational justification, let me briefly describe the manifestations of those characteristics in the area where I have previously encountered them.

The first characteristic is one I have recently been calling "local holism." Many of the referring terms of at least scientific languages cannot be acquired or defined one at a time but must instead be learned in clusters. In the learning process, furthermore, an essential role is played by explicit or implicit generalizations about the members of the taxonomic categories into which those terms divide the world. The Newtonian terms 'force' and 'mass' provide the simplest sort of example. One cannot learn how to use either one without simultaneously learning how to use the other. Nor can this part of the language acquisition process go forward without resort to Newton's second law of motion. Only with its aid can one learn how to pick out Newtonian forces and masses, how to attach the corresponding terms to nature.

From this holistic acquisition procedure a second characteristic of

4. The most explicit and developed formulations are recent: T. S. Kuhn, "What Are Scientific Revolutions?" Occasional Paper 18, Center for Cognitive Science (Cambridge, MA: Massachusetts Institute of Technology, 1981), reprinted in *The Probabilistic Revolution*, vol. 1, *Ideas in History*, ed. L. Krüger, L. J. Daston, and M. Heidelberger (Cambridge, MA: MIT Press, 1987), pp. 7–22; also reprinted in this volume as essay 1; "Commensurability, Comparability, Communicability," in *PSA 1982: Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association*, vol. 2, ed. P. D. Asquith and T. Nickles (East Lansing, MI: Philosophy of Science Association, 1983), pp. 669–88; reprinted in this volume as essay 2. For what I now take to be an implicit, but perhaps more sophisticated, version of the same themes, see my far older paper,

scientific languages follows. Once acquired, the member terms of an interrelated set can be used to formulate infinitely many new generalizations, all of them contingent. But some of the original generalizations or others compounded from them prove to be necessary. Look again at Newtonian force and mass. The force of gravity might have been inverse cube rather than inverse square; Hooke might have discovered that the restoring force of elasticity was proportional to the square of the displacement. These laws were fully contingent. But no imaginable experiment could change merely the form of Newton's second law. If the second law failed, replacing it with another would result also in a local alteration of the language in which Newton's laws had previously been stated. Conversely, the Newtonian terms 'force' and 'mass' can function successfully only in a world in which Newton's second law holds.

I have called the second law necessary, but the sense in which that is so needs further specification. In two respects, the law is not a tautology. First, neither 'force' nor 'mass' is independently available for use in a definition of the other. In any case, the second law, unlike a tautology, can be tested. One can, that is, measure Newtonian force and mass, insert the result in the second law, and discover that the law fails. Nevertheless, I take the second law to be necessary in the following language-relative sense: if the law fails, the Newtonian terms in its statement are shown not to refer. No substitute for the second law is compatible with Newtonian language. One can use the relevant parts of the language unproblematically only so long as one is committed to the law. For this situation, the term 'necessary' is perhaps inappropriate, but I have no better. 'Analytic' clearly will not do.

Return now to the near-trivial justification of the norms or desiderata for theory choice, and begin by asking about the people who embody those norms. What is it to be a scientist? What does the term 'scientist' mean? The word itself was coined around 1840 by William Whewell. What evoked it was the emergence, beginning at the end of the previous century, of the modern use of the term 'science' to label a still-forming set of disciplines that were to be set beside and contrasted with such other disciplinary clusters as those labeled 'fine arts', 'medicine', 'law', 'engineering', 'philosophy', and 'theology'.

Few or none of these disciplinary clusters can be characterized by a

"A Function for Thought Experiments," reprinted in *The Essential Tension: Selected Studies in Scientific Tradition and Change* (Chicago: University of Chicago Press, 1977), pp. 240–65.

set of necessary and sufficient conditions for membership. Instead, one recognizes a group's activity as scientific (or artistic, or medical) in part by its resemblance to other fields in the same cluster and in part by its difference from the activities belonging to other disciplinary clusters. To learn to use the term 'science', one must therefore learn also to use some other disciplinary terms like 'art', 'engineering', 'medicine', 'philosophy', and perhaps 'theology'. And what thereafter makes possible the identification of a given activity as science (or art or medicine, etc.) is its position within the acquired semantic field that also contains these other disciplines. To know that position among the disciplines is to know what the term 'science' means or, equivalently, what a science is.

The names of disciplines thus label taxonomic categories, several of which must, like the terms 'mass' and 'force', be learned together. That local linguistic holism was the first of the characteristics isolated above, and, again, a second characteristic goes with it. The terms that name the disciplines function effectively only in a world that possesses disciplines quite like our own. To say, for example, that in Hellenic antiquity science and philosophy were one is to say also, and paradoxically, that in Greece before the death of Aristotle there was no enterprise quite classifiable as philosophy or as science. The modern disciplines have, of course, evolved from ancient ones, but not one for one, not each from an ancient progenitor appropriately viewed as a (perhaps more primitive) form of the same thing. The actual progenitors require description in their own terms, not in ours, and that task calls for a vocabulary that divides up, categorizes, intellectual activities in a way different from our own. Finding and disseminating a vocabulary that permits description and understanding of older times or of other cultures is central to what historians and anthropologists do.[5] Anthropologists who refuse the challenge are called "ethnocentric"; historians who refuse it are called "Whig."

This thesis—the need for other languages to describe other times and cultures—again has a converse. While we speak our own language,

5. The force of this point depends critically on the claim, developed and defended in "Commensurability, Comparability, Communicability," that the language required to describe some aspects of the past (or another culture) is not translatable into the native language of the person who provides the description. I have provided an extended example of the difficulties created by forcing a modern disciplinary taxonomy on the past in my "Mathematical versus Experimental Traditions in the Development of Physical Science," reprinted in *The Essential Tension*, pp. 31–65.

any activity that we label 'science', or 'philosophy' or 'art', and so on, must necessarily display pretty much the same characteristics as the activities to which we customarily apply those terms. Just as access to Newton's second law is required in order to pick out Newtonian forces and masses, so picking out the referents of the modern vocabulary of disciplines requires access to a semantic field that clusters activities with respect to such dimensions as accuracy, beauty, predictive power, normativeness, generality, and so on. Though a given sample of activity can be referred to under many descriptions, only those cast in this vocabulary of disciplinary characteristics permit its identification as, say, science; for that vocabulary alone can locate the activity close to other scientific disciplines and at a distance from disciplines other than science. That position, in turn, is a necessary property of all referents of the modern term 'science'.

Of course a science need not possess all the characteristics (positive or negative) that prove useful in identifying disciplines as sciences: not all sciences are predictive; not all are experimental. Nor need it always be possible, using these characteristics, to decide whether a given activity is science or not: that question need not have an answer. But a speaker of the relevant disciplinary language may not, on pain of self-contradiction, utter statements like the following: "The science *X* is *less* accurate than the non-science *Y;* otherwise the two occupy the same position with respect to all disciplinary characteristics." Statements of that sort place the person who makes them outside of his or her language community. Persistence in them results in communication breakdown and, if elaborated, often in charges of irrationality as well. One can no more decide for oneself what 'science' means than what science is.

Now, of course, I am back where I began. The person who named *X* a science, *Y* not, was doing the same thing as the person who, earlier in this paper, preferred *X* to *Y* when both were scientific theories. Both violated some of the semantic rules that enable language to describe the world. An interlocutor who supposed their usage normal would find them guilty of self-contradiction. An interlocutor who recognized their usage as aberrant would be hard put to imagine what they could be trying to say. It is not, however, merely language that these statements violate. The rules involved are not conventions, and the contradiction that results from their abrogation is not the negation of a tautology. Rather, what is being set aside is the empirically derived taxonomy of disciplines, one that is embodied in the vocabulary of disciplines and applied by virtue of the associated field of disciplinary characteristics.

That vocabulary can fail to describe, but not, I have argued, merely term by term. Instead, failure must be met by the simultaneous adjustment of large parts of the disciplinary vocabulary. And until the adjustment has occurred, the person who preferred X to Y is simply opting out of the scientific language game. That, I believe, is where the near-trivial approach to justifying norms for theory choice gets its bite.

That bite is, of course, limited. Hempel is right to point out that the near-trivial approach provides no solution to the problem of induction. But the two now do make contact. Like 'mass' and 'force', or 'science' and 'art', 'rationality' and 'justification' are interdefined terms. One requisite for either is conforming to the constraints of logic, and I have made use of it to show that the usual norms for theory choice are justified ("rationally justified" was redundant). Another requisite is conforming to the constraints of experience in the absence of good reasons to the contrary. Both display part of what it is to be rational. One does not know what a person who denies the rationality of learning from experience (or denies that conclusions based upon it are justified) is trying to say. But all that simply provides background for the problem of induction, which, viewed from the perspective developed here, acknowledges that we have no rational alternative to learning from experience, and asks why that should be the case. It asks, that is, not for a justification of learning from experience, but for an explanation of the viability of the whole language game that involves 'induction' and underpins the form of life we live.

To that question I attempt no answer, but I would like one. Together with most of you, I share Hume's itch. Preparing this paper has made me realize that the itch may be intrinsic to the game, but I am not ready for that conclusion.