[9] Although this passage was copied in 1478, its exact date of origin is difficult to pinpoint. Other manuscripts from this collection are believed to have existed since before the fourth century in one form or another (Stillman, 1960).

[10] However, this is an unusual (perhaps a transitional) account of the elements. The elements listed are earth (or metal), water, couperose (or sulfate), and fire, with air not explicitly mentioned.

[11] There were several minor variants of this system of correspondences (e.g., Crosland, 1978, p. 80).

[12] An alternative way of describing the alchemical aesthetic would be to say that the relations involved are extremely nonspecific: for example, "associated with by some path." Under that description, the alchemists would not be guilty of shifting relations between parallel analogs. However, this degree of nonspecificity of relations would still constitute a marked difference from modern scientific usage.

[13] For one thing, it is not clear that the alchemists' analogies are so much less accessible than modern analogies. To the extent that alchemical correspondences were based on surface similarity, they could often be readily guessed. In contrast, in modern scientific analogy the object correspondences are often impossible to grasp without a knowledge of the domain theory, since they are based purely on like roles in the matching relational system.

[14] It might be better to say "rediscovered," since the Greeks, including Plato and Aristotle, used analogy in the modern way.

[15] Alchemy continued into the eighteenth century and beyond, but with greatly decreased influence.

[16] This is apart from variation in the degree to which individuals in our culture conform to our ideal of rationality, as opposed to relying on superstitions based on metaphor and metonymy.

21

*Metaphor and theory change: What is "metaphor" a metaphor for?

RICHARD BOYD

Introduction

In the now classic essay "Metaphor" (Black, 1962b), Max Black considers and rejects various formulations of the "substitution view" of metaphor, according to which every metaphorical statement is equivalent to a (perhaps more awkward, or less decorative) literal statement. Black devotes most of his critical attention to a special case of the substitution view, the "comparison view," according to which a metaphor consists in the presentation of an underlying analogy or similarity. It is clear from Black's discussion that he understands the comparison view as entailing that every metaphorical statement be equivalent to one in which some quite definite respect of similarity or analogy is presented, and that successful communication via metaphor involves the hearer understanding the same respect(s) of similarity or analogy as the speaker.

Black argues that, except perhaps in cases of *catachresis* – the use of metaphor to remedy gaps in vocabulary – the comparison view is inadequate. As an alternative, Black proposed the adoption of an "interaction view" of metaphor. According to this view, metaphors work by applying to the principal (literal) subject of the metaphor a system of "associated implications" characteristic of the metaphorical secondary subject. These implications are typically provided by the received "commonplaces" about the secondary subject. Although Black's position has many facets, it is clear that, at a minimum, it differs from the comparison view in denying that the success of a metaphor rests on its success in conveying to the listener or reader some quite definite respects of similarity or analogy between the principal and secondary subjects: metaphors are, on Black's view, more open-ended (this is not his terminology) than the comparison view would suggest.

In certain passages, Black appears to suggest even stronger points of divergence between his view and the comparison account. In addition to denying that successful metaphors must convey to the reader or hearer some quite definite respect of similarity or analogy, Black also denies that any analysis of an interaction metaphor in terms of explicit analogies or similarities, however elaborate, can capture the cognitive content which it is capable of conveying (Black, 1962b, p. 46). Black sees these features of metaphor as indicative of an important difference between metaphorical uses of language and those uses which have the features of explicitness characteristic of scientific usage.

> We need the metaphors in just those cases where there can be no question as yet of the precision of scientific statements. Metaphorical statement is not a substitute for formal comparison or any other kind of literal statement but has its own distinctive capacities and achievements. (p. 37)

In particular, in this view, one should expect that when metaphorical language is employed in a scientific context, its function should either lie in the pretheoretical (prescientific?) stages of the development of a discipline, or in the case of more mature sciences, it should lie in the realm of heuristics, pedagogy, or informal exegesis, rather than in the realm of the actual articulation or development of theories.

In formulating my own views about the role of metaphor in theory change, I have found it valuable to compare and contrast my understanding of scientific metaphors with Black's account of metaphors in general. Roughly speaking, what I should like to argue here is this: There exists an important class of metaphors which play a role in the development and articulation of theories in relatively mature sciences. Their function is a sort of *catachresis* – that is, they are used to introduce theoretical terminology where none previously existed. Nevertheless, they possess several (though not all) of the characteristics which Black attributes to interaction metaphors; in particular, their success does not depend on their conveying quite specific respects of similarity or analogy. Indeed, their users are typically unable to precisely specify the relevant respects of similarity or analogy, and the utility of these metaphors in theory change crucially depends upon this open-endedness.

On the other hand, I shall argue, this particular sort of open-endedness or inexplicitness does not distinguish these metaphors from more typical cases of scientific terminology, nor need it be the case that these metaphors forever resist complete explication of the relevant respects of similarity and analogy; such explication is often an eventual consequence of successful scientific research. There are, I shall argue, cases in which complete explication is impossible, but far from being indications of the imprecision of metaphorical language in science, such cases reflect the necessity of obtaining a *precise* fit between scientific language and a messy and complex world. The impression that metaphors must lack the precision characteristic of scientific statements reflects, I shall argue, an extremely plausible but mistaken understanding of precision in science.

More precisely, what I shall argue is that the use of metaphor is one of many devices available to the scientific community to accomplish the task of *accommodation of language to the causal structure of the world.* By this I mean the task of introducing terminology, and modifying usage of existing terminology, so that linguistic categories are available which describe the causally and explanatorily significant features of the world. Roughly speaking, this is the task of arranging our language so that our linguistic categories "cut the world at its joints"; the "joint" metaphor is misleading only in that it obscures the fact that the relevant notion of "joint" may be context, or discipline, relative. An important special case of this task of accommodation (not under that description) has recently been investigated in some detail by Kripke (1972) and Putnam (1975a, 1975b), who have emphasized the ostensive character of some of the mechanisms by which the reference of certain natural-kind terms is fixed. Although their accounts of the "causal theory of reference" differ in important details, each of them emphasizes that it is by virtue of the ostensive character of these reference-fixing mechanisms that it is possible for natural-kind terms to refer to kinds which are determined by explanatory or "real" essences, rather than by definitional or "nominal" essences: that is, their account explains how it is possible for natural-kind terms to play a role in what I am calling *accommodation.* The accounts which Putnam and Kripke offer are particularly well suited to cases in which the reference of a term is specified by display of one or more examples of a substance whose real essence is its internal constitution. What I shall argue here is that the employment of metaphor serves as a nondefinitional mode of reference fixing which is especially well suited to the introduction of terms referring to kinds whose real essences consist of complex relational properties, rather than features of internal constitution.

If I am right, this conclusion provides the basis for a clearer understanding of the (itself metaphorical) notion that reference fixing in the case of theoretical terms in science involves *ostension.* I shall suggest that the notion of ostension, and indeed the notion of reference itself, are fundamentally epistemological notions, and that the issue of reference for a general term is the issue of its role in making possible socially coordinated *epistemic access* to a particular sort of thing or natural phenomenon. In terms of the notion of epistemic access, one can formulate an account of linguistic precision in science which allows an adequate treatment of ostensively defined terms in general and metaphor in particular.

I shall also argue that an understanding of certain uses of metaphor in science can teach us something about the nature of the "real essences" which define the natural kinds to which scientific terms refer. The cases of natural kinds defined by real essences on which Putnam and Kripke chiefly rely as examples are cases in which the real definition of the kind in question is (at least on a suitable idealization) provided by a set of necessary and sufficient conditions – conditions whose conceptual representation might become our concept of the relevant kind if our scientific investigations are sufficiently successful. Thus the examples of a posteriori real definitions of natural kinds on which the literature has largely been built suggest that the classical empiricist picture according to which a kind is properly defined by a set of necessary and sufficient conditions united by the mind is basically correct as an idealization. The error of empiricist philosophers, it would seem, was their failure to recognize that such conceptual unities are properly subject to an external and nonconventional requirement of accommodation to appropriate causal structures.

I shall argue that an emphasis on such cases draws our attention away from an important fact about the scientific task of accommodating our language and concepts to the causal structure of the world and that attention to the role of metaphors in science can help us to see how this is so. In particular, I shall argue that for a class of scientifically important kinds – *homeostatic property cluster kinds* – the conception of kind definitions as conceptually unified necessary and sufficient conditions fails even as an idealization. The definitions of such kinds, I shall argue, differ in three important ways from the definitions envisioned by empiricists. In the first place, the properties (relations, etc.) that constitute the definition of a homeostatic property cluster kind are united *causally rather than conceptually.* Such a kind is defined by a family of properties that are causally united in nature: there are causal mechanisms ("homeostatic" mechanisms to use the metaphor I prefer) that tend to bring about their co-occurrence. This fact, rather than any idealized conceptual representation of those properties, is what constitutes their unity as elements in a kind-definition.

Second, the properties that define a homeostatic property cluster kind do not, even as an idealization, specify necessary and sufficient conditions for kind membership. Imperfect homeostasis, I shall argue, dictates that we take homeostatic property cluster kinds to have irremediable indeterminacy in extension – indeterminacy which could not be remedied by any more "precise" definition without abandoning the scientifically crucial task of accommodation of kind definitions to actual causal structures.

Finally, the empiricist picture of kinds as defined by conceptually united sets of necessary and sufficient conditions is mistaken in yet another way when applied to homeostatic property cluster kinds. Sets are individuated extensionally so that the members of a set cannot change from time to time or place to place. Homeostatic property cluster definitions, by contrast, are individuated nonextensionally. Mechanisms of property homeostasis are often themselves not static and as they vary over time or from place to place the properties that make up the definition of a single kind may themselves vary. Numerically the same definition may embody different properties at different times (or places) although defining numerically the same kind!

Homeostatic property cluster kinds are relevant to our understanding of metaphors in science in at least four ways. In the first place, the role of theory-constitutive metaphors in science reflects in a perhaps surprising way the epistemological necessity (and hence the methodological necessity) for the accommodation of conceptual structures to the causal structure of the world. Scientific kinds and categories must be defined in ways which reflect a deference to the world even at the cost of conceptual complexity. The fact that scientific investigation sometimes requires reference to kinds whose definitions are necessarily causally rather than conceptually unified indicates the depth of that necessity.

Moreover, the existence of kinds with the sort of definitional complexity that homeostatic property cluster kinds exhibit helps to explain why theory-constitutive metaphors are so stable a feature of the study of complex systems; indeed such metaphors may be especially important for the investigation of homeostatic property cluster phenomena.

An understanding of homeostatic property cluster kinds also, I shall argue, enhances our understanding of technical matters that bear on our understanding of the semantics of theory-constitutive metaphors. In the first place, an understanding of the semantics of scientific metaphors requires a critique of empiricist conceptions of linguistic precision and recognition of the special features of homeostatic property cluster definitions greatly enhances that critique. More important, as I suggested earlier, an understanding of the semantics of theory-constitutive metaphors requires that we understand reference of linguistic expressions in epistemic terms – in terms of relations of socially coordinated *epistemic access* between language users and features of the world. I shall argue that both epistemic access and the more basic phenomenon of knowledge in terms of which it is defined are themselves homeostatic property cluster phenomena. Thus an understanding of such phenomena is central to the task of understanding the epistemic and semantic function of scientific terms generally and theory-constitutive metaphors in particular.

Examples of metaphor in science

There is, no doubt, a considerable variety of sorts of metaphors that play a role in science, and in theory change. Certain metaphors, which might be plausibly termed exegetical or pedagogical metaphors, play a role in the teaching or explication of theories which already admit of entirely adequate nonmetaphorical (or, at any rate, less metaphorical) formulations. I have in mind, for example, talk about "worm-holes" in general relativity, the description of the spatial localization of bound electrons in terms of an "electron cloud," or the description of atoms as "miniature solar systems."

The fact that these metaphors, and others like them, do not convey theoretical insights not otherwise expressible does not indicate that they play no important role in theory change. Kuhn's work has made it clear that the establishment of a fundamentally new theoretical perspective is a matter of persuasion, recruitment, and indoctrination. It cannot be irrelevant to those enterprises that there is a body of exegetically, or pedagogically, effective metaphors.

Nevertheless, it seems to me that the cases of scientific metaphor which are most interesting from the point of view of the philosophy of science (and the philosophy of language generally) are those in which metaphorical expressions constitute, at least for a time, an irreplaceable part of the linguistic machinery of a scientific theory: cases in which there are metaphors which scientists use in expressing theoretical claims for which no adequate literal paraphrase is known. Such metaphors are *constitutive* of the theories they express, rather than merely exegetical. It might seem doubtful that such theory-constitutive metaphors exist; after all, it is at least plausible that metaphorical language is fundamentally pretheoretical, and lacks the explicitness and precision characteristic of scientific theories. Still, if one looks at theory construction in the relatively young sciences like cognitive psychology, one finds theory-constitutive metaphors in abundance. The examples that I know best are metaphors in cognitive psychology that are drawn from the terminology of computer science, information theory, and related disciplines. The following examples are but a small subset of the actual cases:

1. the claim that thought is a kind of "information processing," and that the brain is a sort of "computer";
2. the suggestion that certain motoric or cognitive processes are "preprogrammed";
3. disputes over the issue of the existence of an internal "brain-language" in which "computations" are carried out;
4. the suggestion that certain information is "encoded" or "indexed" in "memory store" by "labeling," whereas other information is "stored" in "images";
5. disputes about the extent to which developmental "stages" are produced by the maturation of new "preprogrammed" "subroutines," as opposed to the acquisition of learned "heuristic routines," or the development of greater "memory storage capacities" or better "information retrieval procedures";
6. the view that learning is an adaptive response of a "self-organizing machine";
7. the view that consciousness is a "feedback" phenomenon.

I do not want to maintain that *all* of these examples are of fundamental importance to theoretical psychology. Nevertheless, the prevalence of computer metaphors shows an important feature of contemporary theoretical psychology: a concern with exploring analogies, or similarities, between men and computational devices has been the most important single factor influencing postbehaviorist cognitive psychology. Even among cognitive psychologists who despair of actual machine simulation of human cognition, computer metaphors have an indispensable role in the formulation and articulation of theoretical positions. These metaphors have provided much of the basic theoretical vocabulary of contemporary psychology (Neisser, 1966; G. A. Miller, 1974).

Moreover, it is clear that these computer metaphors are theory-constitutive: psychologists do not, generally speaking, now know how to offer literal paraphrases which express the same theoretical claims. This is made clearly evident by the current discussion among psychologists and philosophers about the doctrine called "functionalism" (Block & Fodor, 1972; Block 1977; Boyd, 1980; Fodor, 1965, 1968; Lewis, 1971; Putnam, 1967, 1975b, 1975f; Shoemaker, 1975b). It is widely agreed that some version or other of the doctrine that mental and psychological states are functional states of organisms represents the cognitive content of the metaphorical statement that the brain is a sort of computer. But even among psychologists and philosophers who are convinced that functionalism is true, there is profound disagreement about important issues regarding its interpretation. Thus, this metaphor, and other computer metaphors employed in psychological theorizing, share with more typical interaction metaphors, at least for a time, the property that their cognitive content cannot be made explicit.

In important respects, however, these theory-constitutive metaphors are highly atypical. In the first place, they undergo a sort of public articulation and development that is uncharacteristic of literary metaphors. Typically, a literary metaphor has its "home," so to speak, in a specific work of a specific author; when the same metaphor is employed by other authors, a reference to the original employment is often implicit. When the same metaphor is employed often, by a variety of authors, and in a variety of minor variations, it becomes either trite or hackneyed, or it becomes "frozen" into a figure of speech or a new literal expression (see Black on "orange"). Literary interaction metaphors seem to lose their insightfulness through overuse: the invitation to explore the various analogies and similarities between the primary literal subject and the metaphorical secondary subject becomes pointless or trite if repeated too often. Theory-constitutive scientific metaphors, on the other hand, become, when they are successful, the property of the entire scientific community, and variations on them are explored by hundreds of scientific authors without their interactive quality being lost. They are really conceits rather than metaphors – and conceits which extend not through one literary work, but through the work of a generation or more of scientists.

There is another closely related respect in which theory-constitutive metaphors are atypical. As Black points out, the task of offering explicit and literal paraphrases of literary metaphor is the appropriate task of literary critics and other commentators, but represents an enterprise relatively distinct from the production of the literary works in which the metaphors occur: the task of explication of metaphor is typically separate from the task of production and is often pursued by a quite different group of practitioners. In the case of scientific metaphors, on the other hand, especially in the case of theory-constitutive metaphors, this division of tasks and division of labor does not obtain. It is part of the task of scientific theory construction involving metaphors (or any other sort of theoretical terminology) to offer the best possible explication of the terminology employed. Although this task is sometimes also the preoccupation of professional philosophers (as in the case of functionalism), it is certainly the routine responsibility of working scientists. The sciences in general, and psychology in particular, are self-reflective disciplines, and the explication of theoretical concepts – metaphorical or not – is an essential part of the task of scientific inquiry.

Finally, whatever the merits of the claim that the cognitive content of literary metaphors can never be captured by literal paraphrase, there seems to be no reason to doubt that such explication is possible in the case of some theory-constitutive metaphors, nor is there any reason to doubt that complete explications are often the eventual result of the attempts at explication which are central to scientific inquiry. (For some reservations about explication of theory-constitutive metaphors for cases involving homeostatic property cluster kinds see the section in this chapter entitled "Homeostasis, reference, and precision.")

Literary interaction metaphors, it would seem, display what might be termed *conceptual open-endedness:* they work by inviting the reader (or hearer) to consider the principal subject of the metaphor in the light of associated implications – typically – of the commonplace conception of the secondary subject. Even in those cases in which the metaphor depends upon esoteric information about the secondary subject, the information is of the sort the sufficiently sophisticated reader might be expected to possess (sophisticated commonplaces, so to speak); indeed, the whole point of most literary metaphors would be lost if this sort of knowledge on the part of readers could not be presupposed. The function of literary metaphor is not typically to send the informed reader out on a research project.

Exactly the opposite is the case with theory-constitutive metaphors. They display what might be called *inductive open-endedness.* Although the intelligibility of theory-constitutive metaphors rests on the reader's being able to apply to her current understanding of the primary subject some of the associated implications appropriate to her current conception of the secondary subject, the function of the metaphor is much broader. The reader is invited to explore the similarities and analogies between features of the primary and secondary subjects, including features not yet discovered, or not yet fully understood. This programmatic research-orienting feature of theory-constitutive metaphors explains, I believe, the ways in which such metaphors both resemble and differ from ordinary interaction metaphors. Theory-constitutive metaphors are introduced when there is (or seems to be) good reason to believe that there are theoretically important respects of similarity or analogy between the literal subjects of the metaphors and their secondary subjects. The function of such metaphors is to put us on the track of these respects of similarity or analogy; indeed, the metaphorical terms in such metaphors may best be understood as referring to features of the world delineated in terms of those – perhaps as yet undiscovered – similarities and analogies. Thus, it is hardly surprising that, at least for a time, it is not known exactly what the relevant respects of similarity or analogy are; many have yet to be discovered or understood. Similarly, it is unsurprising that theory-constitutive metaphors can retain their interactive quality even though they are employed, in a number of variations, by a number of authors, and over a long time. Repeated employment of such metaphors does not consist (as it would in the case of more typical interaction metaphors) of merely repetitive and trite invitations to once again explore the same understanding of the principal subject in the light of the same body of associated implications about the secondary subject. Instead, the use of theory-constitutive metaphors encourages the discovery of new features of the primary and secondary subjects, and new understanding of theoretically relevant respects of similarity, or analogy, between them.

Precisely because theory-constitutive metaphors are invitations to future research, and because that research is aimed at uncovering the theoretically important similarities between the primary and secondary subjects of the metaphors, the explication of these similarities and analogies is the routine business of scientific researchers, rather than of some specialized body of commentators. Indeed, the explication of such metaphors is essentially an automatic consequence of success in the research programs that they invitė. For this reason, and because we cannot know a priori that such investigations will not ultimately be completely successful with respect to the issues raised by any particular metaphor, we have no reason to deny that complete explication of theory-constitutive metaphors is sometimes possible. They do, however, share with literary metaphors the important property that their utility does not depend on the (even tacit) availability of such an explication. Indeed, the utility of theory-constitutive metaphors seems to lie largely in the fact that they provide a way to introduce terminology for features of the world whose existence seems probable, but many of whose fundamental properties have yet to be discovered. Theory-constitutive metaphors, in other words, represent one strategy for the accommodation of language to as yet undiscovered causal features of the world.

Accommodation and reference fixing

The possibility that linguistic usage in science might be accommodated to as yet unknown "joints" in the causal structure of the world is at least as old as the seventeenth century. In Book III (and, to a lesser extent in Books II and IV) of the *Essay Concerning Human Understanding* (Locke 1690/1959), Locke explores the consequences of the proposal that substance terms should be taken to refer to the "real essences" of substances in the sense suggested by Boyle's corpuscular theory of matter: that is, that terms like "gold" or "water" should be understood to refer to specific kinds of corpuscular structure. Locke's rejection of this proposal in favor of the view that general terms must refer to kinds specified by "nominal essences," that is, by criteria of membership fixed by definitional convention, has formed the basis for all subsequent empiricist discussions of meaning and reference.

It is important to remember that empiricist accounts of meaning and reference, from Locke to the present, have rested on what are essentially "verificationist" principles. It is insisted that the kinds referred to by general terms must be delineated by "nominal essences" precisely because attempts to delineate kinds by "real essences" or "secret powers" (Hume's phrase) would make knowledgeable use of language impossible. Thus, for example, Locke rejects the possibility of classifying substances on the basis of their atomic structure on the grounds that the limitations of the senses precludes our ever discovering atomic structures. He also concludes that, in general, our inability to discern the hidden inner constitution of things makes knowledge of general laws impossible (although, like all empiricists, he restrains such skepticism where it suits his own philosophical purposes).

More recent empiricists have departed from both Hume and Locke in holding that verificationism *is* compatible with the view that knowledge of general laws is possible. Nevertheless, contemporary versions of Locke's account of general terms – doctrines to the effect that the reference or cognitive content of general terms must be fixed by "operational definitions," "criterial attributes," "law-clusters," "reduction sentences," and the like – still have a verificationist foundation. We are to understand the kinds referred to by general terms to be specified by definitional conventions because – so the argument goes – knowledge of unobservable underlying "powers" or "inner constitutions" is impossible.

In recent years there has been a movement in the philosophy of science away from the instrumentalism implicit in this sort of position and toward the "scientific realist" position that knowledge of "unobservables" and of causal powers is indeed possible (Boyd, 1973, 1983, 1985a, Putnam, 1975a, 1975b, 1975c, Smart, 1963). To a large extent, this realist tendency has resulted from analyses of the actual findings and methods of the empirical sciences. More recently, it has also been associated with the development of distinctly non-Humean positions in epistemology: causal theories of perception, causal theories of knowledge generally, the increased realization that Locke and Hume were correct in insisting that inductive generalization is unfounded unless our categories correspond to causal powers or natural necessity, and the recognition that knowledge of general laws in science is – typically, at least – impossible without some knowledge of unobservable entities or powers (Boyd, 1973, 1983, 1985a, 1985b, 1990a; Goldman, 1967, 1976; Harré & Madden, 1975; Shoemaker, 1975a).

One of the most interesting aspects of this realist tendency has been the recently renewed interest in alternatives to the standard empiricist accounts of language. It has long been recognized that considerations of circularity preclude *all* general terms possessing reference-fixing descriptive definitions. Logical positivists, for example, typically assumed that general terms for qualities of sense-data (e.g., "orange" as a description of the qualitative character of sensation) and their reference (or extension) are fixed ostensively, that is, by association with examples of the relevant sensory qualities, rather than by verbal definition. A number of philosophers, most notably Kripke and Putnam, have recently defended and developed the view – which has occasionally tempted certain of the more realist logical empiricists like Feigl – that the reference of natural-kind terms (like "water" and "gold"), and of theoretical terms in science, might be fixed "causally" or ostensively rather than by definitional convention. Such a view accords with the realistic position that knowledge of "unobservable" causal powers and constituents of matter is possible, and with Quinean dicta to the effect that there are no analytic definitions, no truths by convention. These dicta, in turn, have been confirmed by the experience of philosophers of science, who have found it extraordinarily difficult to find rationally defensible grounds for deciding which measurement procedures are operational definitions or which laws are in law-clusters or which statements are reduction sentences: that is, by the experience of philosophers who have found it impossible in practice to distinguish scientific truths-by-convention from high level empirical truths.

Putnam's account of ostensive reference for theoretical terms (Putnam, 1975a, 1975b) is perhaps the most widely discussed (although, as he points out, it represents a preliminary discussion). The examples of ostensive reference fixing on which he concentrates are those in which terms are introduced that refer to substances like water, or to fundamental physical magnitudes like electrical charge. In these cases, on his view, we may think of the reference of terms like "water" or "electrical charge" as being fixed by "dubbing" or naming ceremonies (the latter terminology is Putnam's, the former, Kripke's) involving the association of the term in question with a sample or exemplary causal effect of its referent, or with a description of stereotypical samples or causal effects. One might, for example, imagine a ceremony in which someone says "Let's call 'water' whatever substance is present in this bucket over here" or "Let's call 'electrical charge' whatever fundamental physical magnitude is responsible for the deflection of the needles on meters of this sort." Of course, no general terms are actually introduced precisely by such ceremonies, but as an idealization such an account does indicate how unambiguous (or nearly unambiguous) reference could be achieved without explicit definition in terms of necessary-and-sufficient conditions, criterial attributes, or operational procedures. If the stuff in the bucket is nearly pure water, then we may take the dubbing to have fixed the reference of "water" as the chemical substance water, even though no one at the time of the dubbing may know what property is really essential to water (i.e., no one might know that water is H_2O and that ice and steam are, therefore, species of water). Similarly, if electrical charge is the only one fundamental magnitude which is the principal determinant of the position of needles in the meters in question, then we might think of the reference of "electrical charge" as having been fixed as electrical charge – even if no one knew how to detect electrical charge generally, or what its most fundamental properties are, or how, in other circumstances, to distinguish the effects of charge from the effects of other fundamental magnitudes. This sort of ostensive reference fixing, then, can be understood as a procedure aimed at accommodation of linguistic usage to as yet undiscovered causal structures: we introduce terminology for substances and fundamental magnitudes by appealing to situations in which we believe they are exemplified, prior to our discovery of their fundamental or essential features (that is, prior to the discovery of those properties which would have to be mentioned in an extensionally correct explicit definition).

The success of the particular style of ostensive reference fixing by citation of samples or exemplary effects also depends on the particular sorts of kinds at which the ostension "aims." It is possible to employ samples in fixing the reference of substance terms precisely because relatively pure samples of substances are possible – because it is possible to have situations in which only one kind of the sort in question is significantly present. However, in cases where terms are to be introduced ostensively to refer to kinds whose essential properties include causal relations to, and co-occurrence with, other kinds of the same sort, one must typically think of their reference as being fixed by practices with a much more complicated structure than the dubbings considered by Putnam. The avoidance of quite fundamental ambiguity requires that one think of the mechanisms of reference fixing as involving features which serve to disambiguate the references of each introduced term between two or more kinds of the same sort, all of which are manifested in the exemplary samples, or situations. There must, therefore, be differences between the ways in which the different terms are introduced corresponding to the differences in essential properties between the kinds referred to. The rationale for ostensive introduction of general terms is to permit reference to kinds whose essential properties may not yet be known – and thus to accommodate linguistic categories to as yet only partially understood causal features of the world. It is thus typically impossible that the differences between the essential properties of such co-occurring kinds should be marked, either in actual reference fixing or in idealized models of dubbing, by entirely accurate and complete descriptions of their respective essential properties.

If I am right, one of the important roles of theory-constitutive metaphors is to accomplish nondefinitional reference fixing of this sort. If the fundamental properties of the metaphorical secondary subjects of a body of related metaphors are sufficiently well understood, then these metaphors can be employed – together perhaps with exemplary circumstances of application – to fix nonliteral referents for the metaphorical expressions they contain. If the differences in essential properties of the secondary subjects are sufficiently great, analogy or similarity with them may suffice to disambiguate the (new) reference of the terms thus introduced. Subsequent metaphors which develop the same metaphorical theme may be used to report discoveries or theoretical speculations regarding the kinds to which the new reference of the metaphorical terms so introduced has been fixed. In such cases, as in the case of ostensive introduction of terms for substances, linguistic terminology is accommodated to the structure of natural phenomena whose fundamental features are not yet fully understood.

The case of computer metaphors in cognitive psychology, I believe, illustrates this sort of ostensive introduction of theoretical terminology. Mental and psychological states and processes are, almost certainly, among the sorts of kinds whose essential properties are relational – they are functional states or processes. Furthermore, it seems reasonable to hold about many psychological states and processes that their causal relations to *other psychological states and processes* are among their essential properties. (These claims make up what is least controversial about the doctrine called "functionalism.") Theoretical terms in psychology, then, are among those for which reference fixing must typically involve disambiguation between several quite different but co-occurring kinds of the same sort. Computational states and processes of the sort which are the secondary metaphorical subjects of computer metaphors in psychology are also functional states and processes: typically their essential properties are their causal relations to other computational states or processes or to the inputs and outputs of the machines which realize or manifest them. What I am suggesting is that – when computer metaphors in cognitive psychology are successful – the metaphorically employed computer terms come to have new referents in the context of psychological theory construction. They refer to functionally defined psychological states or processes which bear to each other functional relations analogous to those which the literal referents of these terms bear to one another. If the metaphors are apt, and if they are drawn in sufficient detail, the differences in functional (relational) properties of the literal referents of the computer terms will serve – by analogy – to disambiguate the referents of these terms in their theory-constitutive metaphorical applications.

The example of computer metaphors has several features that illustrate the programmatic character of theory-constitutive metaphors and, indeed, of nondefinitionally introduced theoretical terminology generally. In the first place, when we inquire about the referents of theory-constitutive metaphorical expressions, it is necessary to inquire about the aptness or insightfulness of the metaphors in which they are employed. The introduction of theory-constitutive metaphors, like the introduction of any theoretical terminology, represents an estimate that natural phenomena of the right sorts exist (see Putnam, 1975a, pp. 224–5). Computer metaphors are introduced into psychological theory on the basis of an informed "guess" that there are important similarities or analogies between their primary and secondary subjects. The aim of the introduction of such terminology is to *initiate investigation* of the primary subjects in the light of an informed estimate of their properties. In cases where a theoretical metaphor proves not to represent a real insight, we need no more inquire about the new referents of its metaphorical terms than we do with respect to the referent of the term "vital force": in such cases the "guess" does not work out, and the relevant terms do not refer at all. If there are no features of human cognition closely analogous to the fundamental features of machine computation, then, for example, there is nothing which is the referent of "information processing" in its use as a term of psychological theory.

It is also possible to see an important relation between the programmatic inductive open-endedness of theory-constitutive metaphors and their role as reference-fixing devices. Computer metaphors are introduced in order to make possible investigation of the similarities and analogies between human cognition and machine computation. These metaphors are (at least for a while) open-ended precisely because the research program they help to initiate is incomplete: we still do not know in exactly what respects human cognition resembles machine computation. What metaphorical uses of computer terminology permit is that we introduce at a relatively early stage theoretical terms to refer to various plausibly postulated computerlike aspects of human cognition, which then become the objects of further investigation. Two features of this sort of use of metaphors are worth remarking on here. First, it is by no means necessary – in order that fundamental ambiguities among the new referents of the metaphorical terms be avoided – that the fundamental or essential properties of their literal (nonmetaphorical) referents be fully understood. The relation between "indexing" of memory items and subsequent "information retrieval" may not be fully understood by computer specialists, for example, but this does not prevent the notions of "indexing" and "retrieval" from playing a role in successful psychological metaphor. Indeed, this sort of case illustrates the important point that successful theory-constitutive metaphors can have programmatic importance for the development of research about their metaphorical secondary subjects, as well as for research about their primary subjects and the new referents of their metaphorical terms. There is no doubt that analogies between human cognition and computer functioning have provided useful heuristics for computer scientists as well as for psychologists.

The second point is that subsequent developments in the research program initiated by the employment of theory-constitutive metaphors may well lead to further articulation of the metaphor in question, to the introduction of new theoretical terminology (metaphorical or not) and to a consequent refinement of usage and further reduction in ambiguity of such terminology. Thus, for example, the notion of "feedback," in its information-processing and computer usage, seems to cover a variety of sorts of phenomena, ranging from simple feedback circuits in analogue devices to "system monitoring" functions in complex computational devices. If there is some insight in the claim that consciousness is a sort of feedback, then one might expect that, as both applied computation theory and psychology progress, new terminology will be introduced (perhaps, but not necessarily, by employment of further computer metaphors), which will permit the drawing of finer distinctions among those cognitive processes that are analogous to the machine processes now grouped under the term "feedback."

These programmatic features of theory-constitutive metaphors – the fact that they introduce the terminology for future theory construction, refer to as yet only partially understood natural phenomena, and are capable of further refinement and disambiguation as a consequence of new discoveries – explain the fact that repeated employment and articulation of these metaphors may result in an increase in their cognitive utility rather than in a decline to the level of cliché.

What is significant is that these programmatic features of theory-constitutive metaphorical expressions are, in fact, typical of theoretical terms in science (in fact, they are true of general terms, generally). Normally, we introduce terminology to refer to presumed kinds of natural phenomena long before our study of them has progressed to the point where we can specify for them the sort of defining conditions that the positivist's account of language would require and, indeed, where no such conditions may exist at all (see the discussion of homeostatic property cluster kinds below). The introduction of theoretical terms does require, however, some tentative or preliminary indication of the properties of the presumed kinds in question. Any such terminology must possess a sort of programmatic open-endedness, inasmuch as its introduction fixes a presumed topic for future research. Thus the introduction of theoretical terminology generally requires just the features that theory-constitutive metaphors provide. One way of providing a tentative and preliminary account of the properties of presumed kinds – and of disambiguating terms referring to presumed kinds of the same general sort – is by open-ended analogy to kinds whose properties are in some respects better understood. One way of expressing such analogies is by metaphorical use of terms referring to those better understood kinds. Theory-constitutive metaphors, then, simply represent one strategy among many for the preliminary stages of theory construction.

Given the initial plausibility of the view that metaphorical uses of language are insufficiently precise to be scientific, it is surprising that the use of metaphors in science is so unsurprising. The fact of their utility in science is philosophically important, not because they represent an especially unusual phenomenon, but, instead, because they provide an especially apt illustration of ubiquitous but important features of scientific language generally.

Metaphor and linguistic precision: Challenges for a theory of reference

Black argues, and common sense concurs, that metaphorical language lacks the precision of scientific language. Against this eminently plausible position, I have proposed that there exist theory-constitutive metaphors in abundance, and that a nondefinitional "causal" or "ostensive" account of reference of the sort advanced by Kripke and Putnam can be employed to defend the view that the metaphorical terms occurring in theory-constitutive metaphors actually refer to natural kinds, properties, magnitudes, and so on – hereafter referred to simply as "kinds" – which constitute the nonliteral scientific subject matter of such metaphors. I suggested that, in fact, the use of theory-constitutive metaphors represents a nondefinitional reference-fixing strategy especially apt for avoiding certain sorts of ambiguity.

Is this the whole story? Is the intriguing issue of metaphor and theory construction, in reality, reducible to a footnote to an already extant theory of scientific language? Is it a mistake to believe that important theoretical questions are raised by the issue of metaphor in scientific theory construction?

That the answer to these questions is no is suggested not only by common sense, but also by the following consideration: nothing in the application of the causal theory of reference to theory-constitutive metaphors directly addresses Black's claim that metaphorical language lacks scientific precision. Supposing it is established that there are theory-constitutive metaphors, and that their metaphorical terms are to be understood referentially, the question still remains why their apparent imprecision does not render them unsuitable for scientific theory construction. No alternative to the common-sense understanding of linguistic imprecision has been offered. Furthermore, this deficiency – which the issue of theory-constitutive metaphors forces us to examine – reflects, not a defect in my presentation, but rather a serious limitation in the existing accounts of the causal theory of reference and, as I shall argue toward the end of this essay, in our understanding of the nature of the definitions of some scientifically important kinds.

There are, broadly speaking, two rival accounts of the ways in which reference is fixed for natural-kind terms and the other sorts of general terms that occur in scientific and everyday discourse. According to the empiricist account, for all but a special class of primitive terms, all general terms are to be understood as governed by stipulatory definitional conventions. Certain sentences (operational definitions, law-clusters, meaning conventions) involving such terms are true by stipulation, are known a priori, and fix the meaning (and the extension or reference) of the terms they define. According to the rival causal or ostensive accounts, for many general terms, reference is fixed by an appropriate sort of causal interaction between users of the term and instances of the kind to which it refers.

The empiricist account readily supplies criteria of linguistic precision. Two uses of the same term (or of two lexicographically different terms for that matter) are coreferential or co-extensive only when they are governed by the same definitional conventions. Vagueness arises from inexplicitness or intersubjective variation in definitional conventions, and ambiguity from the association of a single term with two or more nonequivalent definitional conventions. Both sources of linguistic imprecision have the same remedy: each general term should be associated with a single, quite explicit, and definite conventional definition which is accepted by the relevant linguistic community *prior to* the employment of the term in question. Linguistic precision can be identified with the existence of explicit, detailed, and intersubjectively accepted conventional definitions.

The empiricist account of general terms has other important consequences as well. It entails that there are certain (definitional) statements which are immune from revision or refutation by experimental evidence. It entails that these definitional truths can be established by convention prior to the conduct of experimental investigations. Futhermore, it entails that necessary truths are almost always a priori definitional truths, and (as Kuhn, 1970a, brilliantly observed, especially pages 101–2) it entails that major changes in scientific theories are almost always to be diagnosed as changes in subject matter or conceptual framework, rather than as new discoveries. Finally, although the empiricist account of general terms may be construed as a theory of *reference,* in important respects it represents a nonrealist, nonreferential account of general terms. The extensions of the "natural" kinds, or values of the magnitudes which are the referents of general terms are – according to the empiricist conception – largely fixed by arbitrary and empirically unrevisable definitional conventions. General terms are not understood as referring to independently existing kinds or magnitudes. Indeed, an antirealist and verificationist perspective has provided the defense of the empiricist account of language since its proposal by Locke (see Locke, 1690/1959, especially Book III). Characteristically, the empiricist account of language treats even precisely defined nonobservational ("theoretical") general terms as playing a merely heuristic, or conceptual (but nonreferential), role in scientific theory construction.

With the decline of logical empiricism, especially within the philosophy of science, each of these consequences of the empiricist account of language has come to seem unacceptable to a number of philosophers. It is the unacceptability of these consequences that has led philosophers like Kripke and Putnam (and, much earlier, Feigl – see Feigl, 1956) to advance causal or ostensive theories of reference. The reasoning goes something like this: The empiricist account of language has unacceptable philosophical consequences; on the other hand, if the reference of some general terms is fixed nondefinitionally, in a way somewhat analogous to ostension, baptism, or the employment of stereotypical "definitions," then one can accommodate the variety of antiempiricist findings of recent philosophy of science and philosophy of language. Thus, a nondefinitional account of reference fixing is probably correct for a wide variety of general terms.

There is nothing wrong with this sort of reasoning. Indeed, the proposal that reference-fixing mechanisms are typically nondefinitional promises to be one of the most important achievements of recent analytical philosophy. What has not happened, however, is the articulation of a genuine causal *theory* of reference as an alternative to the received empiricist theory. It is proposed that reference is somehow a nondefinitional causal relation, but no general theory of nondefinitional reference in the literature integrates all of the proposed nondefinitional reference-fixing strategies into a single unifying theory of reference. In particular, existing proposals lack an adequate account of ambiguity and coreferentiality, and, thus provide no account either of the nature of linguistic precision, or of the methodological or linguistic practices which are apt for achieving it. The causal account of reference arises from an attempt to defend the position that general terms (especially "theoretical" terms in scientific theories) should typically be understood referentially, that general terms can refer even though they do not possess unrevisable conventional definitions, and (what is important) that tokens of a term employed in different contexts, at different historical times, within different paradigms, or in different "possible worlds" may be coreferential, even though they are not associated with equivalent conventional definitions. The independent philosophical justification for these doctrines warrants the acceptance of a causal account of reference, but it remains true that no available account offers a satisfactory treatment of the crucial issues of coreferentiality, ambiguity, and linguistic precision. In particular, one must reject the "obvious" criterion for coreferentiality of ostensively introduced terms: that two tokens of a term are coreferential only when they are each connected (by a historical causal chain of speakers' intentions to corefer) to the same dubbing or introducing ceremony. There are almost never actual events which can be identified with idealized introducing ceremonies, and, furthermore, painful human linguistic experience makes it abundantly clear that good intentions are not sufficient to avoid unintended ambiguity or shifts in reference.

Arguably, the absence of an adequate account of coreferentiality does not seriously undermine the cogency of the considerations which favor a causal account of reference. The causal account does avoid unacceptable consequences of the empiricist theory of language, and the judgments about reference and about coreferentiality which the causal account protects – even if they are not assimilated into a general theory of reference – are quite well justified by independent philosophical, linguistic, and historical considerations. On the other hand, if we are concerned about the role of theory-constitutive metaphors in science, the situation is reversed. Both common sense and the best available treatments of metaphor suggest that metaphorical language must be imprecise, nonreferential, and essentially heuristic, just as the empiricist theory of language would suggest. In the previous section, I showed that it is *possible* to maintain that the metaphorical terms in theory-constitutive metaphors refer, even though they lack explicit definitions, by adopting a nondefinitional account of the way in which they refer. I have not, however, shown that such metaphorical terms *must* be understood referentially and, in particular, I have not replied to the plausible rebuttal that – precisely because their imprecision precludes their sustaining a definite reference over time, and from one occasion of use to another – metaphorical terms in science should be understood *nonreferentially,* and scientific metaphors should be seen as playing a largely heuristic role in theory construction. Existing causal theories of reference do not provide the machinery for a reply to this challenge.

The considerations that have persuaded many recent philosophers of science to abandon the empiricist (and especially the verificationist) position that theoretical terms play a heuristic or conceptual but nonreferential role in scientific theories are quite general and, if sound, should apply to almost all cases of theoretical language in successful scientific theory construction (see, for example, Boyd, 1973, 1983, 1990a, 1991; Byerly & Lazara, 1973; Feigl, 1956; Fodor, 1968; MacCorquodale & Meehl, 1948; Putnam, 1975a, 1975b; Smart, 1963). If considerations of linguistic precision should dictate a nonreferential and heuristic treatment of theory-constitutive metaphors, serious questions would be raised at least about the generality of the currently accepted antiempiricist account of scientific theory construction. If the considerations that support a referential treatment of theoretical terms, and that support the rejection of Kuhn's paradigm relativism regarding the meaning and reference of such terms, are sound then they should be applicable as well to metaphorical terms in science, and it should be possible to extend them to a general theory of reference which adequately treats the issue of coreferentiality and precision. I shall address the following questions which reflect the challenge we have just examined:

1. Given that it is *possible* to employ a nondefinitional account of reference fixing as an alternative to both the verificationist accounts of empiricists, and the related relativist accounts of Kuhn (1970a) and Hanson (1958), why is such a position preferable to a nonreferential treatment according to which the role of theoretical terms is purely heuristic or conceptual?

2. How do the considerations that constitute the answer to 1 apply to the especially difficult case in which the theoretical terms in question are metaphorical terms occurring in theory-constitutive metaphors?

3. Given that ambiguity and linguistic imprecision are real possibilities in the use of scientific language, what account of ambiguity, coreferentiality, and linguistic precision can the defenders of nondefinitional reference fixing offer as an alternative to the received empiricist account?

4. How does this alternative account treat the especially difficult issue of imprecison in theory-constitutive metaphors?

In order to answer these questions, it will be necessary to digress in order to develop, at least in outline, a general theory of nondefinitional reference and a nondefinitional alternative to the received empiricist account of coreferentiality, ambiguity, and linguistic precision. This digression will, I believe, prove fruitful. Scientific metaphors raise truly fundamental issues about language and linguistic competence, and the theory of reference required to understand them has several quite startling consequences, which are important both to an understanding of metaphorical language, and to an understanding of language in general. We shall discover, for example, that there is, in an important sense, no such thing as *linguistic* precision; there are rational strategies for avoiding referential ambiguity, but they are not a reflection of rules of linguistic usage (as the empiricist theory suggests). Rather, they reflect essentially nonlinguistic principles of rational inquiry. We shall also discover that a nonreferential but heuristic treatment of metaphorical language in science is ruled out (as are similar paradigm-relativistic treatments of theoretical terminology of the sort advocated by Kuhn) by quite general epistemological considerations.

Epistemic access: The basis of a theory of reference

Let us then begin our digression into the philosophy of language by addressing the general question: what is reference? What relation between the use of terminology and features of the world is at issue when the question of reference is raised? What sorts of phenomena is a theory of reference supposed to explain? One way to approach the issue of the nature of reference is to examine some of the doctrines about reference which have produced so much recent excitement in philosophy. It is clear that we sometimes refer by pointing, or by employing definite descriptions. If what Putnam and Kripke say is basically correct, then we can also refer to things by employing terminology which bears the right sort of historical relation to antecedent introduction ceremonies, or by employing stereotypical descriptions which look very much like definite descriptions, but are not. In his provocative papers, Field (1973, 1974) suggests that there is a relation of *partial denotation* between certain words and features of the world which is importantly like reference. Putnam (1975e) argues that a *principle of benefit of doubt* is appropriate when assessing the reference of terms in the work of previous scientists, and that a *division of linguistic labor* involving deference to scientists and other experts is essential to reference (Putnam, 1975a).

If these doctrines are even approximately correct (and I believe that they are) then the phenomenon of reference has some quite striking properties: it can be manifested by pointing, by explicitly defining, by dubbing, or by stereotyping; it is essentially connected to the knowledge-gathering efforts of experts and specialists; and it admits of partial manifestation. It is reasonable to ask what sort of relation between language use and the world it is, that has such varied manifestations. Indeed, it is reasonable to ask what the justification is for the presumption that there is a *single* phenomenon of reference with all these different manifestations.

So far as I have been able to ascertain, this question has almost never been explicitly addressed in the literature on reference. The one place in which I have found it treated is Putnam (1975a), where it is suggested that reference and truth should be "so construed that, at least in the 'paradigm case', at least for important classes of sentences, at least if things go as they should, sentences will tend to be accepted in the long run if and only if they are true." This standard for theories of reference and truth is seen as consonant with the "scientific realism" in epistemology defended in Boyd (1973). This principle is, as Putnam remarks, especially suited to explaining the principle of benefit of the doubt and other doctrines which link the notion of reference to issues regarding the opinions and investigations of experts.

The account of reference offered here (which was developed independently of the position of Putnam, 1975a) has Putnam's position as a special case, and may be viewed both as an explanation and a justification for Putnam's position, and as an extension of it which can provide a theoretical basis for a wider variety of recent discoveries about reference. I shall not attempt here to provide an analytic definition of reference, or to establish necessary and sufficient conditions for a word's referring to a particular thing or kind. On Quinean grounds, I doubt that such analytic definitions or specifications of necessary and sufficient conditions are ever to be found in the case of philosophically important concepts and, in any event, I shall argue that for reference no such necessary and sufficient conditions – analytic or otherwise – exist. Furthermore (following Field), I believe that it is a misleading idealization to portray the referential relation between language and the world as being constituted by relations of determinate reference between words and their unique referents. What I shall do is to try to describe the essential features of reference in such a way as to illuminate as much as possible the issues raised by discussions of reference in the recent philosophical literature. I shall be especially concerned to learn from, and to explicate, the grain of truth in each of the following doctrines:

1. Operationalism;
2. The law-cluster account of "meaning" and reference for theoretical terms;
3. Putnam's (1975a) claim that there is a "division of linguistic labor" involved in reference fixing;
4. The suggestion of Quine and Ullian (1970) that language "extends the senses";
5. Gibson's (1966) claim that perception is detection of "ambient information";
6. Putnam's principle of benefit of the doubt;
7. Putnam's suggestion that those who introduced general terms like "water" intended to name an explanatory real essence if there was one;
8. Field's claim that there is a relation of "partial denotation" which may obtain between a general term and more than one kind of natural phenomenon at a time when the distinction between those kinds has not yet been drawn;
9. Causal theories of knowledge and perception (Goldman, 1967, 1976) and the suggestion that they are closely related to causal theories of reference (especially the view that the general reliability of belief-producing mechanisms or methods is a crucial feature of knowledge);
10. The suggestion (Feigl, 1956; Byerly & Lazara, 1973) of a causal theory of detection and measurement for physical magnitudes analogous to causal theories of perception;
11. The view that a realist account of scientific theories (i.e., one which treats theoretical terms as referring to real kinds) is essential to a satisfactory account of the epistemology and method of science. Here we are especially concerned with the view that "theoretical" considerations are essential to the reliable assessment of evidence in science, that the reliability of such theoretical considerations rests on the approximate truth of the body of "collateral theories" upon which they depend, and that rational scientific practice, when successful, eventuates in the adoption of successively more accurate approximations to the truth. (For a defense of this realist and cumulative account of method with respect to issues of measurement, see Cronbach & Meehl, 1956; for a similar treatment of ontological issues in science, see MacCorquodale & Meehl, 1948; for a realist treatment of principles of experimental design and assessment of experimental evidence, see Boyd, 1973, 1983, 1985a, 1985b.)
12. The suggestion of Goldstein (1978) that the metaphor of "ostension" in nondefinitional accounts of reference fixing for theoretical terms is to be understood in terms of the role of those terms in "pointing out" or indicating directions for future research programs.

It seems to me that the grains of truth in these doctrines can best be explained by an account of the essence of reference which generalizes the doctrine of Quine and Ullian that language extends the senses, the doctrine of Gibson that the senses are detectors of ambient information, and the doctrine of Feigl that "verifying evidence is to be viewed as causally related to the evidence's 'theoretical' entities" (Feigl, 1956, p. 17). In the light of these considerations, I propose to defend the following:

1. The notion of reference is fundamentally an epistemological notion. *Semantic Theory – insofar as it is a branch of Philosophy – is a branch of epistemology.*
2. The central task of a theory of reference is to explain the role of language in the acquisition, assessment, improvement, and communication of knowledge, especially the role of language in making possible social cooperation and rational deliberation within these activities. *What is to be explained is our (collective) capacity to successfully detect and discover facts about the world.*
3. The causal theory of reference is true not, primarily, because reference involves causal connections to dubbing or introducing ceremonies, but rather because the referential connection between a term and its referent is typically sustained by a variety of epistemically relevant causal connections both between users of the term and examples of its referent (measurement, perception, detection, experimental manipulation, etc.) and between different users (reporting, deliberating, justifying, disputing, etc.). *A causal theory of reference is true precisely because reference is an epistemological notion and causal theory of knowledge is true.*
4. In deciding issues in the theory of reference it is, therefore, appropriate to make use of the best available epistemological theories. *The true theory of reference will be a special case of the true theory of knowledge: the true theory of reference for theoretical terms in science will be a special case of the true theory of the epistemology of science.*

It also seems to me essential that one adopt a *dynamic* and *dialectical* conception of reference, in contrast to conceptions of reference which present a *synchronic, piecemeal,* and *nondialectical* idealization of the relation between individual words and features of the world. I intend to criticize conceptions of reference according to which the referential relation between a natural language and the world is entirely constituted by those relations of reference which obtain between particular words, and quite definite kinds.

One consequence of such accounts is to treat as nonreferential those connections between words and features of the world which (like Field's partial denotation) do not link words to unique referents. Equally important is the consequence that diachronic changes in linguistic usage, which alter relations of definite reference, are not themselves constitutive of the phenomenon of reference; they must be diagnosed as "changes of reference" in a sense which necessarily contrasts with "report of new discovery." One of the consequences of this sort of picture of reference is the plausibility of the empiricist doctrine that the definiteness and constancy of reference must be guaranteed by explicit and purely conventional definitions for all nonprimitive general terms. The remaining three claims are in opposition to this static conception of reference.

5. The accommodation of our language and conceptual categories to the causal structure of the world ("cutting the world at its joints") is essential in order that knowledge be possible. Since – in the absence of perfect causal knowledge – such accommodation cannot be accomplished by explicit and conventional definitions, nondefinitional procedures for accommodating language to the world are essential to knowledge. Since knowledge gathering is the essential core of reference, *the processes of linguistic accommodation are essential components of reference.*

6. Ostensive reference fixing, and other nondefinitional reference-fixing mechanisms – in the absence of perfect knowledge – will often establish referential connections between a word and more than one thing or kind. Routinely, terms with this sort of "imprecision" play a vital role in the socially coordinated discovery and communication of knowledge; indeed, the employment of terms of this sort appears to be essential to scientific inquiry (and rational inquiry generally). Thus, if reference is the relation between language and the world which explains the role of language in the acquisition and communication of knowledge, *nondeterminate referential connections between words and features of the world are essential components of reference.*

7. It is also routine that the acquisition of new knowledge, and the exploration of new areas of inquiry, require that linguistic usage be modified so as to mark newly discovered causal features of the world. This sort of dialectical modification of langauge use (which has what Field, 1973, calls "denotational refinement" as a special case) is essential to the process of accommodation of language to (newly discovered features of) the causal structure of the world, and is thus an essential component of reference. *Reference has an essential dynamic and dialectical aspect. Changes in language use – when they reflect the dialectics of accommodation – do not represent changes of reference in any philosophically puzzling sense of that term. Instead, such dialectical changes of reference are characteristic of referential continuity and represent perfectly ordinary vehicles for the reporting of new discoveries.*

In order to defend this epistemological account of reference, I propose to analyze the notion of reference in terms of the notion of *epistemic access.* I hold that, for any particular general term, the question of reference is to be understood as the question: to which kind (or kinds), or property (or properties), or magnitude (or magnitudes), . . . and so on, does our use of this term afford us epistemic access? When we conduct rational inquiry intended to discover facts about the referent of this term, about what kind(s) do we in fact gather information? A better picture of the relevant notion of epistemic access or information gathering can be obtained by first considering very simple cases of language use which illustrate Quine and Ullian's claim that language extends the senses. Consider, for example, the case of cries issued by sparrows to warn others of approaching predators. Such crying "extends the senses" in a perfectly straightforward sense. Sparrows, hearing such cries, are able to detect indirectly the presence of predators outside their line of sight through the efforts of others. The detection of predators takes on a social character: sparrows have, in such cases, socially coordinated epistemic access to certain kinds of predators. Even though it may be inappropriate to talk of "reference" in such cases, a "warning cry" is a warning cry rather than a mating call precisely because sparrows (a) can detect predators by sight with fair reliability, (b) typically issue warning cries only when they do so, and (c) typically respond to hearing warning cries in much the same way that they respond to seeing a predator.

The sort of epistemic access afforded by certain words in human languages (like "red" or "cold," for example) is quite analogous. Central to our employment of the term "red" are the facts that most speakers of English can detect the presence or absence of the color red, use the term "red" to report the presence of that color, and (under normal circumstances) take others' reports of "red" as indicative of the presence of the color red.

To these simple cases of language extending the senses, we may add cases in which the relevant "detection" skills are cognitive rather than merely perceptual (consider the case of the general term "refrigerator"), and where "discovery" rather than "detection" may be the more appropriate term (but not, I insist, "construction"). More relevant to the issue of reference for theoretical terms in science are cases in which epistemic access, the discovery of facts about the referents of terms, requires scientific investigation and serious theory construction. In such cases, the role of general terms in the social coordination of knowledge acquisition is substantially more complex. It remains true in these cases that one may be afforded a passive extension, if not of one's senses, then of one's research. I am, for example, able to know that DNA carries the genetic code in mammals, by relying on the testimony of experts whose research demonstrated this fact about the referent of "DNA." Inasmuch as this is the only way I can obtain such information about the substance which is the referent of "DNA," however, the sort of epistemic access which I have to its referent is not central to determining the reference of "DNA."

What is important is the epistemic access which the term "DNA" affords to DNA in virtue of the role that term plays in the organization of research. Here there are at least three distinguishable ways in which use of the term "DNA" makes it possible for the relevant scientific community to make of itself an instrument for the detection (or discovery, if you prefer) of information about DNA:

1. Its use permits scientists to report to each other the results of studies of DNA.
2. Its use permits the public articulation, justification, criticism, debate, and refinement (in the light of justification, criticism, debate, *and* experimentation) of theories concerning DNA, thus making the interpretation of data and the evaluation of proposed theories – as well as the reporting of results – into a social enterprise.
3. Finally, the use of the term "DNA" makes possible verbal reasoning concerning DNA with respect to questions of data interpretation, theory evaluation, experimental design, and so forth. That is, the use of language makes possible not merely the formulation of theories and publicity and cooperation in their assessment; it makes it possible for reasoning (whether individual or public) to be *verbal* reasoning: to take place in words.

In discussing simpler cases of epistemic access, it was necessary to appeal to the general and typical reliability of the human senses, or the commonplace cognitive ability to, for example, recognize refrigerators. In the context of theoretical terminology, the analogous factors are somewhat more complex. In the first place, of course, the epistemic reliability which is involved is (typically) that of the community of scientific experts, rather than that of particular individuals, especially laymen. Furthermore, the scope of the relevant notion of epistemic reliability must be formulated with some circumspection. Scientific terms must be understood as providing the sort of epistemic access appropriate to the level of epistemic success typical of scientific discoveries. Historical evidence suggests that the theories which are accepted by the scientific community are rarely entirely correct in every respect, even when they reflect the discovery of fundamentally important truths. What is typical of successful scientific investigations is successive improvements in partial but significant knowledge: scientific progress typically arises from the replacement of revealing (though only approximately accurate) theories with more revealing (and more nearly accurate) theories. Similarly, it is true that the history of science reveals a number of plausible but fundamentally mistaken "false starts" which are only corrected over time (for example, Darwin's belief in inheritance of acquired characteristics, or the theory of vital forces). Thus the sort of success which is characteristic of epistemic access in the case of a theoretical term in science involves the capacity of the scientific community, typically and over time, to acquire increasingly accurate knowledge about the referent of that term.

The mark of reference, then, is epistemic access, and the mark of epistemic access is the relevant sort of socially coordinated epistemic success. Roughly, a general term, *T*, affords epistemic access to a kind (species, magnitude, and so on), *k*, to the extent that the sorts of considerations which are (in the relevant historical context) rationally taken as evidence for statements involving *T* are, typically, indicative in an appropriate way of features of *k*. The following mutually supporting epistemic relations between a term, *T*, and a kind, *k*, are characteristic (but by no means definitive) of the sorts of relations which constitute epistemic access:

1. Certain of the circumstances or procedures which are understood to be apt for the perception, detection, or measurement of *T* are, in fact, typically apt for the perception, detection, or measurement of *k*.[1]

2. Some of the circumstances which are taken to be indicative of certain features or properties of manifestations of *T* are, in fact, typically indicative of those features or properties of manifestations of *k*.

3. Certain significant effects attributed to the referent of *T* by experts (or generally, in the case of nontheoretical terms) are in fact typically produced by *k*.

4. Some of the most central laws involving the term *T* are approximately true if they are understood to be about *k*.

5. There is some generally accepted, putative, definite description of the referent of *T* which is in fact true of *k* and of no other kind.

6. The sorts of considerations which rationally lead to modifications of, or additions to, existing theories involving the term *T* are, typically and over time, indicative of respects in which those theories can be modified so as to provide more nearly accurate descriptions, when the term *T* is understood as referring to *k*, so that the tendency over time is for rationally conducted inquiry to result in theories involving *T* which are increasingly accurate when understood to be about *k*.

It is, of course, possible for a term, *T*, to afford epistemic access to several quite different kinds. The term "demon" probably afforded epistemic access to a great variety of kinds of natural phenomena for centuries. What I am suggesting is that it is correct to talk of the referent of a general term precisely in those cases in which the term affords substantial epistemic access to a single kind or, at any rate, to a family of closely related ones. The mark of reference is continued epistemic success with respect to information gathering about a particular kind. In the case of general terms employed in the theoretical sciences, such continued success is typically reflected in theoretical advances and in new discoveries, but it is important to realize that the same phenomenon of continued epistemic success is reflected in a more mundane fashion in the case of everyday general terms. Even in the absence of profound discoveries or theoretical advances regarding refrigerators, the color blue, or candlesticks, it remains true that we daily succeed in conveying to each other new and reliable information regarding refrigerators, blue things, and candlesticks, by employing the terms "refrigerator," "blue," and "candlestick."

I want to defend the view that reference is constituted by just this sort of epistemic access, in part, by showing how such a view can make sense of our seemingly incompatible philosophical intuitions regarding reference. But it should be said at the outset that the analysis of reference in terms of epistemic access has considerable independent plausibility. It is hard to see how language could serve the vital social functions it does if epistemic access were not a central feature of its use, and – given the limits of human knowledge – it is hard to see how the relevant sort of epistemic access could be other than that which involves gradual improvement of knowledge. Furthermore, a referential treatment of theoretical terms – and a treatment which explains how reference is possible prior to definitive knowledge – is apparently essential to any adequate treatment of the role of theoretical considerations in the assessment of scientific evidence (Boyd, 1973, 1983, 1985a, 1985b, 1990a). Thus, considerations of both everyday and scientific epistemology favor an account of reference in terms of epistemic access. It remains to show that such an account also makes sense of the received body of philosophical truisms about reference.

In the first place, an account of reference in terms of epistemic access avoids the necessity for idealized reference to dubbing ceremonies and for an implausible emphasis on the role of speakers' referential intentions. The kind to which a general term refers is determined by the role that term plays in socially coordinated inquiry, rather than by any particular features of its introduction, or the intentions of the speakers who first introduced it. It is true, of course, that the history of a term's use, and the intentions of those who use it, will play a role in determining the kind(s) to which it affords epistemic access, but we are able to offer an account of reference which does not make introducing events or speakers' intentions definitive in this regard.

Similarly, we can see how an epistemic-access account of reference accommodates the insights of the two most important logical empiricist accounts of the meaning of theoretical terms: operationalism and the "law-cluster" account. Operationalism insists that the reference (or the cognitive content – many early defenders of operationalism rejected a referential treatment of theoretical terms) is determined by conventionally fixed procedures of detection or measurement. It is mistaken in holding that detection and measurement procedures are definitive in reference fixing, and even more mistaken in holding that the reliability of measurement or detection procedures is a matter of linguistic convention. Nevertheless, if an epistemic-access account of reference is correct, there is an important grain of truth in operationalism. In typical cases, substantial epistemic access to a natural kind or physical magnitude rests on the possession of relatively reliable detection or measurement procedures. Furthermore, the sort of continued theoretical understanding characteristic of reference for theoretical terms will typically result in the acquisition of even more sophisticated and accurate techniques of detection or measurement. This sort of centrality of detection or measurement to referential epistemic access represents the important grain of truth in operationalism.

In the case of the law-cluster account, there are several grains of truth which can be accommodated to the epistemic-access account. In the first place, of course, if continued epistemic success is characteristic of reference, then the intuition is vindicated that there is something absurd in the suggestion that all of our most fundamental beliefs about a "theoretical entity" might be fundamentally mistaken. Substantial and sustained epistemic access guarantees that we cannot be entirely mistaken all the time. Furthermore, inasmuch as scientific investigations tend to be influenced by those theoretical beliefs that the scientific community considers most fundamental, continued epistemic success provides strong indication that those beliefs are, in some respects at any rate, correct. None of these considerations, of course, supports the view that fundamental laws are true by linguistic convention, or that it is always true when reference occurs that most of them are even approximately correct.

The central importance of law-clusters in reference is also explained by a central feature of scientific methodology. Scientific methodology is heavily theory-determined: for example, one tests a proposed theory by trying to identify those alternative theories which are – in the light of the best available theoretical knowledge – most likely to be true, and by designing experiments or observational studies with the aim of choosing between a proposed theory and its *plausible* rivals (this is a point which Putnam has often emphasized). There is every reason to believe that this procedure for theory assessment is crucial to the success of scientific practice (Boyd, 1973, 1983, 1935a). Because the plausibility judgments involved in the practice of this methodological principle depend on applications of the best available theories, the most fundamental laws which the scientific community accepts about a given kind will play a methodologically crucial role in the discovery of new knowledge. Indeed, the approximate truth of such collateral theoretical beliefs is part of the explanation for future epistemic success. We have thus identified another important grain of truth in the law-cluster view: The approximate truth of many of the most important laws in a given subject area is not only a probable consequence of the sustained epistemic-access characteristic of reference, in many cases, it provides the explanation for the success of scientific method in producing the epistemic successes which constitute that epistemic access. Here again, of course, the doctrine that law-clusters are true by stipulation need not be invoked; indeed, that doctrine is contrary to the methodological picture of sustained experimental *and* theoretical criticism and subsequent modification of existing theories.

It is also apparent that the epistemic-access account of reference offers an explanation for many of the less empiricistic doctrines regarding reference and knowledge, which appear on the list with which I began this section. Indeed, the epistemic-access account straightforwardly incorporates causal theories of perception, detection, and knowledge; Quine and Ullian's doctrine that language extends the senses; realist accounts of scientific methodology; and Goldstein's suggestion that ostension involves indication of research directions.

The epistemic-access account also provides an important elaboration of Putnam's talk of a division of linguistic labor. Experts play a crucial role in reference for theoretical terms (and relatively esoteric terms generally) precisely because it is they who provide nonpassive epistemic access to the referents of those terms. In this regard, it is worth remarking that what occurs is not really a division of *linguistic* labor at all. Instead, what is involved is the social division of mental (or, better yet, cognitive) labor: some of us are auto mechanics and know what "accelerator pump" means, others of us are nurserymen and know what "beech" means, whereas still others are physicists who know what "black hole" means. This division of labor is not primarily a linguistic phenomenon, nor is it primarily an epistemological phenomenon: instead, as Putnam insists, it represents facts about social organization of labor at a certain stage of historical development. The division of cognitive labor is related to the issue of reference only because it is reflected in the ways people have of gathering information about features of the world, and because the notion of reference is essentially an epistemic notion.

Consider now the three remaining intuitions about reference on the list with which I began this section: Putnam's plea for "benefit of the doubt" in assessing reference, his claim that the earliest users of general terms like "water" intended to refer to the secret inner constitution of the substance in question, and Field's defense of a notion of partial denotation. If I am right, all three represent commentaries on the dialectical aspects of reference: the accommodation of language to the causal structure of the world.

What I am calling "accommodation" is a response to an epistemological problem, whose discovery represents the principal epistemological achievement of early empiricism: inductive generalization is reliable (or, at any rate, is nonaccidentally reliable) only if the categories in terms of which generalizations are formulated correspond in the right way to the causal

powers of things in the material world (Locke, 1690/1959, Book IV, Chap. iii, Sections 14, 25, 29; Hume, 1739/1973, pp. 90–1). If all *A*'s I have so far examined have produced the effect *B* under circumstances *C*, and I conclude on that basis that *A*'s always produce effect *B* under circumstances *C*, I am going to be right (barring pure luck) only if the categories *A*, *B*, and *C* correspond in the right way to the kinds of causal powers that operated to produce the effects that I observed in the sampled cases. I can go wrong, for example, if all the observed *A*'s actually belong to a smaller kind *D*, and in fact, only *D*'s have the causal power to produce *B* under circumstances *C*; I can go wrong if the category *C* is too broad to capture the causal contribution which the observed instances of *C* made to the observed effect, or if the category *B* is so narrow that it excludes some of the very effects which are sometimes produced by the very mechanisms which were involved in the sample situations. As Locke and Hume (somewhat inconsistently) recognized, it follows that knowledge of general laws is impossible unless we are able to succeed at the "metaphysical" task of "cutting nature at its joints."

It was the fond hope of twentieth-century logical empiricists that a non-"metaphysical" nonrealist account of the meaning of scientific theories would prove to be compatible with an acceptable account of the possibility of their rational confirmation by experiment and observation. The failure of positivist philosophy of science indicates that this was a vain (if well motivated) hope. The emergence of realist conceptions of scientific epistemology reflects the recognition that Locke was right after all: it is impossible to understand scientists as being in the business of achieving nonaccidental success at inductive generalization without understanding them to be in the business of learning about (typically "unobservable") causal powers and underlying mechanisms of structures (Boyd, 1973, 1983, 1985a).

It follows, then, that the business of inductive generalizations must be the business of "cutting the world at its joints" – the business of describing and classifying natural phenomena in ways which in fact correspond to underlying causal powers or mechanisms. In other words, the accommodation of language to the world is essential for linguistically mediated epistemic access. It is for this reason that I insist that the dialectical process of accommodation – the introduction of linguistic terminology, or the modification of current usage so that general terms come to afford epistemic access to causally important features of the world – is an essential component of reference. The sorts of epistemic success that are characteristic of reference are only possible in cases where general terms afford epistemic access to kinds which are "natural" in the sense of corresponding to important causal features of the world.

We are, thus, in a preliminary way, able to offer an explication of Putnam's doctrine that the earliest users of terms like "water" intended to name the underlying real essence, if any, which explained the observable properties of their samples of water. Taken literally, this claim is patently false. What is true, however, is that the earliest users of "water" were embarked on an enterprise – the socially coordinated and linguistically mediated discovery and transmission of information about natural substances – whose rational conduct eventually required the deployment of a general expression which holds of just the inner constitution of the substance which predominated in the samples they called "water." Although they were not in a position to intend, to know, or even to imagine it, the rational conduct of the enterprise upon which they were embarked was to require the employment of terminology coextensive with our term "H_2O."

In a somewhat similar way, we can explicate Putnam's principle of the benefit of the doubt. Suppose that earlier practitioners of some science have achieved a certain measure of success: Suppose, for example, that they have come to be able to make relatively accurate predictions about a significant range of observable phenomena. In the light of the best available accounts of scientific epistemology, we may say that it is overwhelmingly likely (as a partial explanation for their success) that the linguistic terminology of their field afforded them epistemic access to (at least many of) those kinds of natural phenomena that are crucial in the causal determination of the phenomena that they have been able to predict successfully.

Their terminology must have, with some success, "cut the world at its joints." Suppose that we are now in possession of an even more sophisticated theory of the same subject area – an even more sophisticated account of those "joints." If we now ask the question: "to which natural kinds did the terminology of those earlier scientists afford them epistemic access?" we shall (quite properly) answer the question in the light of the best available current theory about what natural kinds causally determine the phenomena at issue. In many typical cases, the most plausible answer will be that their terms were – in all or most cases – coreferential with the same terms as we currently use them. Such an answer – when it is justified – will constitute part of a causal explanation of the epistemic success of earlier researchers – an explanation informed both by historical data and by the best available account of the structure of the world they were studying. Such an account of Putnam's dictum accords perfectly with the claim that what a theory of reference should explain is the role of language in the acquisition, improvement, and communication of knowledge.

We are now in a position to see the wisdom of Field's talk about partial denotation. Consider how accommodation of linguistic usage to the causal structure of the world works in cases where existing practice reflects real errors in classification of natural phenomena. There seem to be two relatively distinct types of error-collecting procedures, which correspond to two different sorts of errors.

In type one cases, we have classified together (say, as *A*'s) certain things which have no important similarity, or we have failed to classify together things which are, in fact, fundamentally similar. In cases of this sort, we typically revise our classifications and say that the things in the first instance really weren't *A*'s after all, and that the things in the second instance really were *A*'s after all.

In type two cases, the situation is more complicated. We have classified things as *A*'s and have met with success in certain sorts of causal generalization or theory construction. We later discover that for certain other theoretical or practical purposes, the things we have classified as *A*'s do not form so natural a kind. Instead, we are led to employ a classification which partly overlaps those cases which we have earlier classified as *A*'s. It may later turn out that one or the other of these two classifications seems the more fundamental, in the sense that it plays a role in the more significant general laws, but it remains true that each of the categories is appropriate for the formulation of interesting generalizations or laws. (I have formulated the description of these two types of errors as though the terms in question referred to natural kinds of things. Obviously, similar cases obtain for properties, kinds of properties, magnitudes, and so on.)

If all cases of classificatory error were of the first type, then the two proposals of Putnam which we have been discussing would be entirely plausible. In such cases, it is plausible to say that we are correct in saying that the erroneous misclassifications represented saying of a non-*A* that it is *A* and saying of an *A* that it is not an *A*. Precisely because there is no other natural kind close to *A* which includes the deviant cases, it is plausible to say that the referent of *A* has remained the same, and that we simply learned more about it. There would be only one kind which might plausibly be thought of as the referent of "*A*," even before the anomalies were discovered, and it would be reasonable both to claim (in the metaphorical sense) that it had been the "intended" referent all along, and therefore to apply the principle of benefit of the doubt and to say that earlier speakers' use of "*A*" had been coreferential with our more sophisticated use. Cases of type one are cases in which no new kind has been discovered; rather they are cases in which the boundary of a previously known kind is fixed with greater accuracy.

Unfortunately, cases of type two are not so clear-cut. Consider the term "fish." At one time, whales and porpoises were classified under the term "fish." Later it was discovered that whales and porpoises are mammals and are quite distinctly unlike other marine vertebrates in many important respects. We now say that whales and porpoises are "not fish" or that they are "not true fish." It remains true, however, that porpoises and bonefish do have many interesting nonphylogenetic features in common. For purposes of many sorts of investigations (of fishing industries, or animal locomotion, for example) it may be perfectly rational to classify whales and porpoises together with the boney and cartilaginous fish.

Suppose that the response to the anatomical, behavioral, physiological, and evolutionary findings that make it rational to distinguish fish from marine mammals had instead been that people had begun to say that there were two importantly different sorts of fish: furry fish and scaled fish (with subsequent modifications of terminology to distinguish "true scaled fish" from, for example, sharks and rays), and that the general term "fish" had continued to be universally applied to bonefish, sharks, rays, porpoises, and whales alike. Suppose, that is, that the important biological discoveries about aquatic animals that we have discussed had not resulted in a change in the animals which people termed "fish," and that the term "fish" had retained its prescientific usage, rather than being employed as a relatively technical term approximately coextensive with "osteichthyes." Under the circumstances we are imagining, people would not have said, for example, "Whales really aren't fish after all," but they would have marked, with different terminology, the same distinctions between kinds of aquatic animals which we now make. Provided that our imaginary linguistic community gets its biological theories right in other respects, it is unreasonable to say that they are mistaken in saying, "Whales are fish." If it is also unreasonable to say that we are wrong in saying, "Whales are not fish," then we have constructed a situation in which Putnam's principle of benefit of the doubt is inapplicable, even though the linguistic communities in question have made no scientific errors. After all, the two communities have the same linguistic history prior to the relevant biological discoveries, and each could – with equal justification – apply the principle of benefit of the doubt to enshrine its own *current* usage as exemplary of the earlier reference of the term "fish."

Two facts are made obvious by examples of this sort. In the first place, the schemes of classification, or modes of measurement, that are inductively appropriate for the acquisition of general knowledge in one field of inquiry may be quite different from those that are appropriate to another. We can and must "cut nature at its joints," but the boundaries between joints are themselves context-specific. Ways of classifying animals which are appropriate for evolutionary biology may be inappropriate for commerce, or for ethological studies. To use Goodman's terminology, "projectability" is a context-of-inquiry relative property of predicates.

Second, when it first becomes evident that it is necessary to draw a distinction between kinds where none has been drawn before, it is often the case that nothing in previous linguistic usage or intellectual practice dictates which of the newly marked kinds, if any, should be referred to by whatever the relevant previously used general term is, and which should be referred to by newly introduced terminology. It is, for example, undetermined whether the old terminology should be co-opted for the more technical, or the less technical, of the subsequent distinctions. In the case of "fish" the term has come to be used in a relatively technical way, whereas in the case of the term "jade," the term has retained its old commercial usage, and new technical terms ("jadite" and "nephrite") were introduced to mark the relevant technical distinction. (This example is from Putnam, 1975a.) As we have seen, neither of these choices was dictated by constraints of rational usage: All that rationality required was that the relevant distinction be marked in the language.

This phenomenon illustrates, I believe, what Field calls partial denotation. It often happens that a term affords epistemic access to two (or more) relatively similar – but clearly distinguishable – kinds during the period before the relevant distinctions have been drawn. Prior to the discoveries that give rise to the drawing of those distinctions, substantial information is gathered about the kinds in question – and formulated with the aid of the general term in question – so that the epistemic access afforded by that term is often crucial to the discoveries in question. After the relevant discoveries have been made, relevant changes in linguistic usage are made, but linguistic and scientific rationality do not dictate a unique new referent for the term in question. These refinements of usage represent the paradigm case of the accommodation of language to the causal structure of the world.

It is clear that the epistemic-access account of reference fully explicates Field's notion of partial denotation. Field is right to think of partial denotation as very closely related to typical cases of reference, because partial denotation involves not only epistemic access but also – in the sorts of cases we are discussing – the sorts of epistemic successes characteristic of reference. Indeed, the eventual resolution of partial denotation in favor of ordinary reference is typically achieved in the light of those successes. It is precisely this sort of "denotational refinement" (Field's term) that one would expect to be commonplace when theory-constitutive metaphorical terms are introduced at early stages of theory construction. As I suggested earlier, the term "feedback" in psychology is a likely candidate.

Field is right in another way. If we think of reference as being the relation between expressions of language and features of the world by virtue of which communication and linguistically mediated discovery are possible, then partial denotation – and indeed epistemic access in general – must be counted as part of that relation between language and the world, and so must the process of linguistic accommodation. There are two reasons why these dialectical elements must be understood as part of the phenomenon of reference itself. First, what is central in reference is epistemic access and epistemic success, and both of these can be achieved to a considerable extent in cases of partial denotation. Second, the accommodation of linguistic categories to the causal structure of the world is essential to the very possibility of the epistemic success characteristic of reference. To think of the phenomenon of reference as excluding these dialectical elements, and as somehow consisting solely of cases in which a term affords epistemic access to only one kind, would be a denial of the basically epistemic character of reference, and might, as well, lead to the absurd conclusion that in early prescientific communities many extremely useful terms have no referential interpretation at all.

There is an even more important consequence of understanding the accommodation of language to the world to be a routine feature of the process of reference, and a routine response to the acquisition of new knowledge. According to the empiricist conception of general terms, there are two quite distinct sorts of changes in the ways we use language. On the one hand, we may use general terms in a way which preserves their current referents, in order to revise, modify, amend, or contradict things we have previously said. On the other hand, we may change our usage in such a way that one or more general terms change their referents (that is, we can "adopt" new criterial attributes, law-clusters, operational definitions, or reduction sentences for such terms). Only the former sort of change, according to an empiricist understanding of reference, represents an appropriate vehicle for the assertion of new discoveries or the refutation of former beliefs. Changes in linguistic practice of the second sort – which involve changes in the referents of the relevant terms – amount, on the empiricist view, to a decision to speak a language which is, with respect to the terms in question, a new language altogether. According to such a view, sentences containing those terms which are uttered before the change in reference are in no straightforward way comparable with sentences containing the same terminology after the linguistic shift: they have a quite different subject matter and their acceptance represents simply the adoption of a new linguistic convention (this is the essence of Kuhn's treatment of the term "mass" during the change from Newtonian mechanics to special relativity; Kuhn 1970a, pp. 101–2).

Against this empiricist relativism, it is possible to insist, as an epistemic-access account of reference does, that law-clusters, operational definitions, and reduction sentences are not definitive of the referent of a general term, are not established by defining conventions, and can be modified or disconfirmed without changing the entity to which a term affords epistemic access (and, thus, without changing its referent if it refers). Nevertheless, cases like those which provide counterexamples to Putnam's principle of benefit of the doubt *are* cases in which it seems impossible to maintain that the relevant term referred to just the same kind before and after the important change in language use.

What are we to say about these cases? Are they cases in which what has occurred is a change in linguistic convention (or "world view," or "conceptual scheme") *and not a discovery?* Or, alternatively, are we – having recognized that Putnam's principle of benefit of the doubt fails in these cases – forced to treat pre-Linnaean uses of "fish" and pre-Einsteinian uses of "mass" nonreferentially on the grounds that there is no unique kind or magnitude (respectively) to which such earlier uses afforded epistemic access? Are we to say that "fish" and "mass" did not refer at all? Or, similarly, are we to say that psychologists' uses of the term "feedback" do not refer because there may be several feedbacklike psychological processes?

According to the view that I am defending here (whether one chooses to say that these terms partially denoted in their earlier usage, or, alternatively, that they referred, but lacked unique referents), the important fact is that they provided substantial and sustained epistemic access to a sufficiently small number of kinds that their use resulted in sustained increases in knowledge (and eventually in the discovery that crucial distinctions had to be drawn between those kinds to which they did afford epistemic access). This sort of linguistically mediated epistemic success – *which necessarily includes modification of linguistic usage to accommodate language to newly discovered causal features of the world* – is the very core of reference. It is just a fact that circumstances arise relatively frequently in which a term affords epistemic access to two or more natural phenomena which are importantly different but which are similar enough in certain respects. Consequently, a considerable amount of theoretically and practically useful knowledge about them can be gathered before the relevant distinctions come to light. The epistemic access provided by such terms plays a crucial role in the acquisition of this information, and in the discovery of the relevant differences; furthermore, such sustained epistemic access is characteristic of reference. Thus, both the relation which such terms bear to the world originally, *and* the modification of usage which accommodates the relevant parts of language more precisely to the causal structure of the world in the light of subsequent discoveries, are central features of the phenomenon of reference.

Two conclusions now follow: in the first place, if we are interested in the "microstructure" of reference – the relations between individual words and the world which go together to constitute the referential relation between language as a whole and the world – then the notions of epistemic access and accommodation are more important than the notion of an individual word's possessing a distinct referent. The situation in which a term affords substantial epistemic access to more than one partial denotation, until more precise accommodation is achieved in the light of later discoveries, is so commonplace that we may think of it as one of the typical ways in which language is connected to the world.

In the second place, we can now see how to answer questions about the distinction between discovery, on the one hand, and the adoption of new linguistic conventions on the other, in cases of accommodation involving partially denoting terms. Contrary to the empiricist account, alterations in the reference (or the partial denotation) of general terms is a perfectly ordinary way of expressing the discovery of new natural kinds. Situations like that of "fish," "jade," and "mass," in which a partially denoting term comes to have a more definite referent as language is accommodated to newly discovered features of the world, are absolutely commonplace. Far from representing the adoption of a new and incomparable language with respect to the terms in question, such developments are marks of referential *success:* these partially denoting terms have facilitated the discovery of new and relevant features of the world.

Having developed an epistemic-access theory of reference, we are now in a position to address the questions regarding coreferentiality, ambiguity, and linguistic precision, which were posed at the beginning of this section and to return to the issue of theory-constitutive metaphors.

Consider first the question of why one should prefer to understand the reference of theoretical terms as continuous during scientific revolutions rather than as changing in ways which make comparison between successive theories impossible. This amounts to the question of why one should apply Putnam's principle of benefit of the doubt. We have just seen in some detail how Putnam's principle, amended in the light of Field's notion of partial denotation, is a consequence of an epistemic-access account of reference. Two points in favor of the continuity account are especially relevant because of their relation to the corresponding question about metaphors. First, according to the epistemic-access account, there are no particular features of the use of a theoretical term (like, for example, a law-cluster, or a particular set of measurement procedures, or a set of reduction sentences) which are conventionally definitive of its referent. Thus, we are not obliged to conclude that the referent of a theoretical term has changed whenever there has been a radical change in the relevant theory, provided that there is some reason to treat the change in theory as a response to additional evidence.

Second, in cases of scientific revolutions, the fact that the subsequent theory resembles the previous one in some important respects provides part of the evidence in support of the latter theory, and this evidential consideration makes sense only on the view that what is involved is the replacement of one approximately accurate and well-confirmed theory by another even better theory *with basically the same subject matter.*

This last point can be put in another way: certainly, in the case of the development of the theory of relativity, the earlier Newtonian theory served a valuable heuristic role in the development of the later theory. Newtonian mechanics provided a valuable guide to the construction of a new theory to account for new and surprising data. What the present account of scientific epistemology dictates is the conclusion that the overwhelmingly likely explanation of the heuristic value of a theory in such a situation is that its terms refer (or partially denote); that it is in important respects true; and that for that reason it can serve as a guide to the formulation of an even more nearly true theory *with relevantly the same subject matter,* that is, one in which most terms preserve their earlier referents

(Boyd 1973, 1989, 1991). The typically positivist move of distinguishing between a theory's being approximately true, and its merely providing a heuristically valuable way of looking at data, fails: in all but contrived and scientifically atypical situations, the only plausible explanation for a theory's heuristic value is that its terms refer, and that it is in some important respects approximately true.

Theory-constitutive metaphors, epistemic access, and referential precision

We can now ask the corresponding question about theory-constitutive metaphors: given that it is possible to employ a nondefinitional account of reference to defend the view that theory-constitutive metaphorical expressions should be understood as referring, why is this view preferable to the view that theory-constitutive metaphorical expressions are nonreferential and are merely heuristically useful?

First, the fact that we are typically unable to provide an explication of theory-constitutive metaphors – that we are typically unable to *define* the relevant respects of similarity or analogy between the primary and secondary subjects of these metaphors – does not, in the light of an epistemic-access account of reference, provide any reason to doubt that the relevant metaphorical expressions refer. The existence of explicit definitions is not characteristic of referring expressions, and is not even a typical accompaniment to sustained epistemic access.

Second, the option of treating theory-constitutive metaphorical expressions as serving a merely heuristic role, rather than treating them referentially, is ruled out by general epistemological considerations from the philosophy of science. If the articulation and refinement of a body of metaphors all involving the same metaphorical theme proves to be genuinely fruitful in scientific theory construction, then the only epistemologically plausible explanation is that most of the relevant metaphorical expressions refer, and that the metaphorical statements in question – when interpreted in the light of the nonstandard referents of their metaphorical terms – express important truths.

Finally, Field's notion of partial denotation, which the epistemic-access account explains, makes it possible to treat metaphorical expressions referentially without ignoring the strong intuition that it is unlikely that such expressions always refer to a single definite kind. The view that the metaphorical terms in successful theory-constitutive metaphors should be understood referentially, but perhaps as affording epistemic access to more than one kind, amounts to an understanding that the epistemic-access account of reference dictates for theoretical terms generally.

At the beginning of the last section, I raised the question of what the standard of coreferentiality was if the empiricist standard of preservation of law-cluster, operational definition, or reduction sentences were not correct. The epistemic-access account provides an answer: reference is continuous if the term in question continues to provide epistemic access to the same kind, or if appropriate episodes of denotational refinement take place. Similarly, several roughly simultaneous uses of the same term are coreferential if they are embodied in patterns of usage which afford epistemic access to the same kind(s).

It must be understood that the issue of epistemic access (and thus of reference) for a particular term is a perfectly ordinary scientific question. One is inquiring about the complex causal relationships between features of the world and the practices of the relevant linguistic community; whether, how, and/or to what extent they, in turn, give rise to the relevant sort of epistemic relations between the term in question and one or more kinds.

In typical cases, one might expect the outcome of such an inquiry to be an independent specification of the kinds to which the term in question affords epistemic access. In the case of theory-constitutive metaphorical expressions, this outcome is precluded (at least until later research makes the explication of the metaphor possible) inasmuch as an independent specification of the relevant kinds would amount to the sort of analysis or explication of which theory-constitutive metaphors typically do not admit. Nevertheless, when we inquire whether a single metaphorical expression occurring in a variety of different theory-constitutive metaphors which develop the same theme, has the same referent (or the same partial denotata) in each of those metaphors, we can have evidence for a positive answer, even though an independent specification of its referent may be impossible. For precisely the epistemological reasons which justify a realist conception of scientific theories, we know that the only plausible explanation for genuinely substantial heuristic value in such an extended series of metaphors is that there is a relatively small number of kinds to which their constituent terms afford epistemic access. Thus, whenever an extended series of related scientific metaphors has genuine scientific value, the overwhelmingly plausible explanation lies in the assumption that each of their constituent metaphorical expressions affords epistemic access to at most a small number of kinds – that is, that reference is constant from one employment of such a metaphorical expression to the other.

Thus, if an epistemic-access account of reference is sound, we have every good reason to hold, in the case of genuinely fruitful theory-constitutive metaphors, that all or most of their constituent metaphorical terms refer and that each of them has the same referent (or approximately the same set of partial denotata) in each of its applications within the relevant theoretical context.

Let us turn now to the question of exactness, or linguistic precision, as it arises in the case of theory-constitutive metaphors. Black holds that metaphors lack the "precision of scientific statements," and that they must

therefore play a role in language different from the role played by the theoretical statements of science. It is evident that it is the open-endedness and inexplicability of interaction metaphors which leads Black to this conclusion. If, as I have argued here, such open-endedness and inexplicitness is typical of theoretical statements and of theoretical terms whose reference is not definitionally fixed, the question arises: what is the right account of linguistic precision in science?

According to the empiricist understanding of scientific terms and scientific method, there are two quite distinct kinds of precision in scientific practice. On the one hand, there is precision in the use of scientific language. Since Locke, the empiricist view has been that this sort of precision is achieved to the extent to which general terms are associated with fixed, conventional, and explicit definitions of their extensions or referents. On the other hand, there is what might be called methodological precision: precision in reasoning, careful experimental design, diligent reporting of data, proper control of experimental variables, precision in measurement, and so forth. The first sort of precision is wholly a matter of the proper following of linguistic rules, whereas the second is a matter of care in treating epistemological issues. Black's insistence that metaphors lack scientific precision must, I believe, stem from a recognition that the use of metaphorical terminology fails to meet the first of these empiricist tests of precision.

Against this view of precision in science, I want to maintain that there is only one sort of scientific precision – methodological, or epistemological precision – and that precision in the use of scientific language is merely one feature, and one consequence, of this methodological precision. There is no purely *linguistic* precision, no mere following of *linguistic* rules, which accounts for precision in the use of theoretical terms.

We may, at the outset, see why the empiricist account of linguistic precision is fundamentally mistaken. The aim of this account is to set standards of linguistic precision which guarantee that each general term will refer to exactly one quite definite kind. The referent (or the extension) of a general term is supposed to be precisely fixed once and for all by conventionally adopted defining criteria. Given human ignorance, it would inevitably be the case (as Locke and Hume recognized) that almost none of the terms introduced according to such pure conventions would correspond to "natural kinds," almost none would "cut nature at its joints." As both Locke and Hume recognized, such terms would in fact prove to be almost useless for the acquisition of any general knowledge whatsoever, and would thus be scientifically useless: indeed, it is hard to see what viable social arrangements could sustain the practice of abiding by such self-defeating conventions.

If the empiricist criterion of precision is a failure, we may still ask what sorts of linguistic difficulties it was designed to avoid. It seems that there are two sorts of linguistic imprecision against which empiricist standards of linguistic precision were directed, although the distinction between them does not seem to have been carefully drawn, even by Locke who may well be the most careful of the empiricists in discussing misuses of language.

On the one hand, there is the difficulty which would arise from idiosyncratic uses of a general term – from circumstances in which someone uses a general term with a different referent or extension from the referent or extension which it has in the idiolects of the typical speakers of his language, or from circumstances in which there was a corresponding sort of mismatch between the uses of a term in two communities which share the same language. On the other hand, there is the sort of difficulty which would arise if the linguistic community as a whole used a general term in an ambiguous or vague way – in a way which left it without any definite referent or extension. Empiricism proposes the same solution – definitional linguistic conventions – to both of these problems.

From the point of view defended here, the first of these problems takes on a somewhat different cast. Because, on the epistemic-access view, reference is a social rather than a private phenomenon, fewer circumstances fit the first of these cases than Locke, say, might have thought. For example, many cases in which an individual has atypical evidential standards for the application of a general term, but in which he also relies on indirect evidence provided by the testimony of others, would be diagnosed as cases in which his use of the term afforded him epistemic access primarily to the same kind to which others referred, but in which his own beliefs about the kind in question were seriously mistaken.

Similarly (and even more obviously), cases in which two communities employed different evidential standards in applications of the same term – standards which were in fact apt for the detection of two different kinds, but in which this fact went unnoticed, so that members of each community relied on reports from members of the other – would often be diagnosed as cases of partial denotation, rather than cases in which definite but different referents could be assigned to the term in the vocabulary of each community. Nevertheless, difficulties of the first sort no doubt do occur, and it is reasonable to inquire what remedy exists for them if an epistemic-access account of reference is correct. Here, it is interesting to note, the remedy involves principles of rational inquiry, which are not distinctly linguistic principles: assess evidence in the light of the best available *generally* accepted theory unless compelling evidence dictates its rejection. Rely on the advice of recognized experts. When your standards of evidence contrast sharply with those of others, seek to identify the source of the conflict, and so on. These are independently justifiable methodological principles, but the consistent application of principles of this sort provides the only rational procedure for uncovering or preventing the difficulties we have been discussing.

The second sort of difficulty arises when the use of a general term affords epistemic access to two or more quite distinct kinds, or (worse yet) to no particular kinds at all. What is to be avoided, then, are situations in which a general term partially denotes rather than refers, or situations in which it affords such diffuse epistemic access that even partial denotation is not achieved. The empiricist solution to this problem is to erect contrived categories as the referents of general terms at the cost of abandoning the project of "cutting nature at its joints." The alternative solution is provided by the ongoing project of continuous accommodation of language to the world in the light of new discoveries about causal powers. Here, the examples of the terms "demon," "fish," "jade," and "mass" are revealing. In each of these instances, a case of quasi-reference or worse was resolved by either a subsequent refinement of usage, or by the abandonment of a term altogether. In each case, the improvement in linguistic usage resulted from new discoveries about the world, rather than from attention to linguistic rules or conventions. In general, this sort of accommodation is achieved by careful and critical research about the structure of causal relations, and in particular by the pursuit of questions like: how reliable are the detection and measurement procedures which we now use? When we take several different procedures to be reliable tests for the same kind, or reliable measures for the same magnitude, are we right in believing that they are all indicative of the same kind or magnitude? How similar are the things we now classify together, and in what respects? How different? What new and undiscovered natural kinds, magnitudes, and so on, must be postulated to account for new data? It is in the *methodologically* precise and diligent pursuit of these scientific questions, rather than in any distinctly *linguistic* practices, that the solution to the problem of diffuse epistemic access lies.

Indeed, all of these questions, as well as the remedies for idiosyncratic usage, are a reflection of a single methodological principle: always inquire, in the light of the best available knowledge, in what ways your current beliefs about the world might plausibly be incomplete, inadequate, or false, and design observations or experiments with the aim of detecting and remedying such possible defects. All of the principles which serve to prevent diffuse epistemic access are special cases of this principle, and there is no application of it which is irrelevant to the dialectical task of accommodation. I conclude, therefore, that there are no distinct principles of *linguistic* precision in science, but rather that linguistic precision is one of the consequences of methodological precision of a quite general sort.

Turning now to the issue of metaphor in science, we can see what realist standards of precision should govern their use. One should employ a metaphor in science only when there is good evidence that an important similarity or analogy exists between its primary and secondary subjects. One should seek to discover more about the relevant similarities or analogies, always considering the possibility that there are no important similarities or analogies, or alternatively, that there are quite distinct similarities for which distinct terminology should be introduced. One should try to discover what the "essential" features of the similarities or analogies are, and one should try to assimilate one's account of them to other theoretical work in the same subject area (that is, one should *attempt* to explicate the metaphor). Such principles of methodological precision are, of course, not importantly different from those that properly govern the use of any sort of theoretical terminology in science, and it is for that reason that we may conclude that the "imprecision" of metaphors does not preclude their employment as constituents of scientific theories.

A final remark about the inexplicitness of theory-constitutive metaphors: theory-constitutive metaphorical terms – when they refer – refer implicitly, in the sense that they do not correspond to explicit definitions of their referents, but instead indicate a research direction toward them. The same thing is apparently true of theoretical terms in science generally.

Now, some thinkers have taken such phenomena as support for an *idealist* conception of scientific understanding, which treats implicit features of scientific knowledge as personal and constructive, rather than as objective and intersubjective. It is beyond the scope of this essay to explore this view in any detail. But it is interesting to reflect that the implicit character of scientific metaphors does not demand any such idealist interpretation. They refer by virtue of *social* and intersubjective (as opposed to personal) mechanisms, which connect scientific research with independently existing ("objective") features of the world.[2] Furthermore, when we understand a theory-constitutive metaphor, there is no reason to believe that we somehow *tacitly* understand the similarities and analogies to which its constituent terms afford epistemic access, just as there is no reason to believe that Newton tacitly understood the Einsteinian account of the referents of his theoretical terminology. These considerations are, of course, not conclusive with respect to an idealist and subjective conception of science; but, taken together with the independent evidence for a realist conception of scientific theories, the fact that an idealist explanation is not required in so obvious an area as that of metaphor in science should give a thoughtful idealist pause.

Metaphors, property homeostasis, and deference to nature

We have seen that there are theory-constitutive metaphors, that a naturalistic epistemic-access account of reference can explain their role in the accommodation of scientific language to the causal structure of the world, and that the same conception of reference can explain why the risk of referential ambiguity associated with such metaphors does not compromise their precision in any scientifically interesting sense of precision. If the role of theory-constitutive metaphors is thus rendered nonmysterious, questions about their role in scientific investigation still remain.

In the first place, there is the question of why metaphors prove so valuable in providing theory-constitutive conceptual frameworks in science. I have already suggested that the introduction of metaphorical terminology reduces the risk (or perhaps the extent) of ambiguity when terms are introduced to refer to functionally or relationally characterized phenomena. Is that it, or are there other ways in which metaphors are especially suited for the introduction of terminology in certain sciences?

There is, moreover, the question of the future of any given theory-constitutive metaphor. I have argued that there is no a priori reason to suppose that a theory-constitutive metaphor will forever escape complete explication, but should we expect that such an explication will typically be the fate of a theory-constitutive metaphor if things go well scientifically. Or should we routinely expect that theory-constitutive metaphors will eventually be abandoned or "frozen"? Suppose, for example, that a metaphorical term is introduced for a chemical compound whose (a posteriori) definition is discovered to be provided by the formula *F*. Should we not expect that after this discovery it will be referred to by *F* rather than by the original metaphorical term, or at any rate that the metaphorical term will become a dead metaphor once the essence of the compound has been discovered and research can be guided by that knowledge? Should this not be the fate of all theory-constitutive metaphors if things go well?

No doubt these questions have quite complex answers, but I think that we can make some headway with them if we examine certain cases of natural kinds and natural kind terms with respect to which our practice of deferring to nature in defining kinds goes somewhat further than the examples of natural definitions like "Water = H_2O" suggest.

The sorts of essential definitions of substances reflected in the currently accepted natural definitions of chemical kinds by molecular formulas (e.g., "water = H_2O") appear to specify necessary and sufficient conditions for membership in the kind in question. Recent *non*naturalistic property-cluster or criterial attribute theories in the "ordinary language" tradition suggest the possibility of definitions which do not provide necessary and sufficient conditions. Instead, some terms are said to be defined by a collection of properties such that the possession of an adequate number of these properties is sufficient for falling within the extension of the term. It is supposed to be a conceptual (and thus an a priori) matter what properties belong in the cluster and which combinations of them are sufficient for falling under the terms. It is usually insisted, however, that the kinds corresponding to such terms are "open textured" so that there is some indeterminacy in extension legitimately associated with property-cluster or criterial attribute definitions. The "imprecision" or "vagueness" of such definitions is seen as a perfectly appropriate feature of ordinary linguistic usage, in contrast to the artificial precision suggested by rigidly formalistic positivist conceptions of proper language use.

I doubt that there are any terms whose definitions actually fit the ordinary language model, because I doubt that there are any significant "conceptual truths" at all. I believe, however, that terms with somewhat similar definitions are commonplace in the special sciences which study complex phenomena. Here is what I think often happens (I formulate the account for monadic property terms; the account is intended to apply in the obvious way to the cases of terms for polyadic relations, magnitudes, etc.):

1. There is a family, *F*, of properties that are contingently clustered in nature in the sense that they co-occur in an important number of cases.

2. Their co-occurrence is, at least typically, the result of what may be metaphorically (sometimes literally) described as a sort of *homeostasis*. Either the presence of some of the properties in *F* tends (under appropriate conditions) to favor the presence of the others, or there are underlying mechanisms or processes which tend to maintain the presence of the properties in *F*, or both.

3. The homeostatic clustering of the properties in *F* is causally important: there are (theoretically or practically) important effects which are produced by a conjoint occurrence of (many of) the properties in *F* together with (some or all of) the underlying mechanisms in question.

4. There is a kind term, *t*, which is applied to things in which the homeostatic clustering of most of the properties in *F* occurs.

5. *t* has an analytic definition; rather all or part of the homeostatic cluster *F* together with some or all of the mechanisms that underlie it provide the natural definition of *t*. The question of just which properties and mechanisms belong in the definition of *t* is an a posteriori question – often a difficult theoretical one.

6. Imperfect homeostatis is nomologically possible or actual: some thing may display some but not all of the properties in *F*; some but not all of the relevant underlying homeostatic mechanisms may be present.

7. In such cases, the relative importance of the various properties in *F* and of the various mechanisms in determining whether the thing falls under *t* – if it can be determined at all – is a theoretical rather than a conceptual issue.

8. Moreover, there will be many cases of extensional vagueness that are not resolvable even given all the relevant facts and all the true theories. There will be things which display some but not all of the properties in *F* (and/or in which some but not all of the relevant homeostatic mechanisms operate) such that no rational considerations dictate whether or not they are to be classed under *t*, assuming that a dichotomous choice is to be made.

9. The causal importance of the homeostatic property cluster *F* together with the relevant underlying homeostatic mechanisms is such that the kind or property denoted by *t* is a natural kind reference important for scientific explanation or for the formulation of successful inductive inferences.

10. No refinement of usage which replaces *t* by a significantly less extensionally vague term will preserve the naturalness of the kind referred to. Any such refinement would either require that we treat as important distinctions that are irrelevant to causal explanation or to induction, or that we ignore similarities that are important in just these ways.

11. The homeostatic property cluster which serves to define *t* is not individuated extensionally. Instead, the property cluster is individuated like a (type or token) historical object or process: certain changes over time (or in space) in the property cluster or in the underlying homeostatic mechanisms preserve the identity of the defining cluster. As a consequence, the properties which determine the conditions for falling under *t* may vary over time (or space), *whereas* t *continues to have the same definition.* This historicity in the way the property cluster definition is individuated is itself dictated by methodological considerations in the disciplines in which *t* is employed: the recognition of the relevant continuities in the historical development of the property cluster is crucial to the inductive and explanatory tasks of those disciplines. Thus the historicity of the individuation conditions for the property cluster is essential for the naturalness of the kind to which *t* refers. I do not envision that this sort of variability in definition will obtain for all of the kinds and kind terms satisfying 1 through 10 and I propose to employ the term "homeostatic property cluster" even in those cases in which 11 fails.

The paradigm cases of natural kinds – biological species – are examples of homeostatic cluster kinds. The appropriateness of any particular biological species for induction and explanation in biology depends on the imperfectly shared and homeostatically related morphological, physiological, and behavioral features that characterize its members. The definitional role of mechanisms of homeostasis is reflected in the role of interbreeding in the modern species concept; for sexually reproducing species, the exchange of genetic material between populations is thought by some evolutionary biologists to be essential to the homeostatic unity of the other properties characteristic of the species and it is thus reflected in the species definition that they propose (see Mayr, 1970). The *necessary* indeterminacy in extension of species terms is a consequence of evolutionary theory, as Darwin observed: speciation depends on the existence of populations that are intermediate between the parent species and the emerging one. Any "refinement" of classification that artificially eliminated the resulting indeterminacy in classification would obscure the central fact about speciation on which the cogency of evolutionary theory depends.

Similarly, the property cluster and homeostatic mechanisms that define a species must be individuated nonextensionally as a processlike historical entity. It is universally recognized that selection for characters that enhance reproductive isolation from related species is a significant factor in phyletic evolution, and it is one that necessarily alters over time the species' defining property cluster and homeostatic mechanisms (Mayr, 1970).

It follows that a consistently developed naturalistic conception of the accommodation of scientific language to the causal structure of the world *predicts* indeterminacy for those natural kind or property terms that refer to complex homeostatic phenomena; such indeterminacy is a necessary consequence of "cutting the world at its joints." Similarly, consistently developed naturalism predicts the existence of nonextensionally individuated definitional clusters for at least some natural kinds, and thus it treats as legitimate vehicles for the growth of approximate knowledge linguistic practices that would, from a more traditional empiricist perspective, look like diachronic inconsistencies in the standards for the application of such natural kind terms.

Homeostatic property cluster definitions represent a special kind of deference to the world: instead of being possible *conceptual phenomena* whose content is dictated by the causal structure of the world, they are themselves naturalistically and (perhaps) historically individuated causal phenomena *in the world.* They provide the most striking examples of the phenomenon of accommodation of scientific language to causal structures. They also provide us with a number of insights into the ways in which theory-constitutive metaphors may contribute to that accommodation.

In the first place, consider the question of why metaphors are so often valuable devices for the introduction of theoretical language. Recall that when a term, *t*, is employed metaphorically as a theoretical term scientists are invited to explore the similarities between the phenomenon referred to by *t* in its metaphorical uses and the phenomenon to which *t* literally refers. The cluster of properties that scientists associate with (real) *t*'s is to guide their thinking about metaphorical *t*'s. If the choice of metaphor is apt this strategy of investigation could be valuable for the study of any sort of phenomenon but, I suggest, it may prove especially valuable in the case in which the phenomenon referred to by *t* in its theory-constitutive metaphorical use is a homeostatic property cluster phenomenon whose essence is given by a property cluster rather than by a set of necessary and sufficient conditions. The metaphor may prove more valuable still if the literal referent of *t* is also a homeostatic property cluster kind whose essential structure is in that respect like that of its metaphorical referent.

I speculate that the latter sort of situation obtains in the case of many famous scientific metaphors like the sustained metaphors of economic competition which have underwritten much of evolutionary theory, the various (human) social metaphors invoked in descriptions of the behavior and ecology of nonhuman animals, and military metaphors in the description of bodily responses to disease.

Consider also the question of whether or not we should expect that theory-constitutive metaphors will in the course of successful science typically become fully explicated or otherwise "frozen." At least insofar as we think of explication of a theory-constitutive metaphor as involving the specification of the natural definition of the phenomenon to which it refers, the case of homeostatic property cluster phenomena suggests a negative answer. Where a theory-constitutive metaphor (or any other expression) refers to such a phenomenon, there is no reason in general to believe that such an explication is even possible. The properties that constitute the homeostatic property cluster may not even be finite in number and they may vary significantly from time to time or place to place, so that a finite characterization of them (much less a cognitively tractable characterization) need not be possible. It is a striking fact that, contrary to what empiricist accounts of scientific language and scientific concepts would suggest, we can refer to and study successfully phenomena that could not possibly have the sorts of definitions empiricists envisioned as essential for scientific investigation.

Homeostasis, reference, and precision

An understanding of homeostatic property cluster definitions can also enhance our understanding of the semantics of theory-constitutive metaphors and other linguistic expressions as well. In offering an epistemic-access account of reference I identified a number of mutually reinforcing factors which, I argued, contributed toward the establishment of a referential connection between a term and a feature of the world. I did not propose to offer a definition of reference in terms of necessary and sufficient conditions for a term to refer to a phenomenon and I suggested that no such conditions exist. I am now in a position to make that claim more precisely. I propose that reference itself is a homeostatic property cluster phenomenon and that the mutually reinforcing factors I identified are some elements of the defining cluster.

I propose, moreover, that reference is a homeostatic property cluster phenomenon precisely because reference is an epistemic phenomenon and knowledge is a homeostatic property cluster phenomenon. All plausible theories of knowledge have it that cases of knowledge differ from other cases of true belief in being appropriately justified, or appropriately reliably produced or regulated, or both. One challenge in epistemology is to specify the degrees and combinations of justification and/or reliability that suffice for knowledge. I suggest that knowledge is in fact defined by a homeostatic cluster of justificatory and reliability producing factors and that this fact explains both the "vagueness" of the notion of knowledge and the failure of efforts to provide (even as an idealization) necessary and sufficient conditions for a true belief to be an instance of knowledge. I develop this theme, without the terminology of homeostatic property cluster phenomena, in Boyd (1983) and I sketch a defense of the related claim that rationality is a homeostatic property cluster phenomenon in Boyd (1990a). I believe that almost all the phenomena of special interest to philosophers are homeostatic property cluster phenomena; for a treatment of moral categories along these lines see Boyd (1988).

An understanding of homeostatic property cluster phenomena also helps to clarify the issue of linguistic precision with respect to scientific terms generally and with respect to theory-constitutive metaphors in particular. We have already seen that with respect to the sort of imprecision that manifests itself as ambiguity in the use of theoretical language the appropriate remedy lies not in seeking a distinctly *linguistic* precision of the sort suggested by empiricists, that is, the adoption of conventional definitions in terms of necessary and sufficient conditions. Instead what is required is the sort of *methodological* precision capable of identifying cases of partial denotation.

Empiricists were concerned as well with a different sort of linguistic imprecision – that displayed by "vague" terminology lacking a determinate extension. Part of the motivation for the empiricist conception of conventional meaning was to provide a remedy for this sort of vagueness. Of course the considerations rehearsed in our earlier discussion of linguistic precision suggest that – where vagueness is a problem to be avoided – the remedy is methodological precision leading to the theoretical resolution of indeterminateness. But, what the homeostatic property cluster conception of some natural kinds indicates is that vagueness in extension is by no means always indicative of any imprecision at all. For some kinds, "vagueness" in the application of the associated terminology is precisely indicative of *precision* in the accommodation of language to the causal structure of the world. If, as I have suggested, many theory-constitutive metaphors refer to homeostatic property cluster phenomena, then we have an additional reason for rejecting the empiricist conception of precision for those cases.

I do not mean to suggest that for such metaphors all the vagueness in their application in practice will correspond to real vagueness in the associated phenomenon. Nor do I suggest that when a literal homeostatic property cluster term is metaphorically used to refer to a homeostatic property cluster phenomenon there will be a neat match between the respects of vagueness of its two referents. I do suggest, however, that the vagueness of scientifically useful theory-constitutive metaphors may serve to remind us of both the actual vagueness of some natural phenomena and the deep limitations of the empiricist conception of linguistic precision.

NOTES

The present chapter is for Herbert Feigl and James J. Gibson. In it I have focused on the question: how can we explain the role of metaphor in the articulation of new scientific theories? I have not addressed the question: what role does metaphorical thinking play in theory invention? I find the second question no less important, but I do not have anything interesting to say about it.

Especially in developing my views on epistemic-access and reference I have benefited from advice and criticism from a number of colleagues. I want especially to thank William Alston, Alex Goldstein, Hilary Kornblith, Barbara Koslowski, Richard Miller, Sydney Shoemaker, Robert Stalnaker, and Nicholas Sturgeon.

The present essay is a revision of the original version which appeared in the 1979 edition of this collection. Apart from minor revisions, the only new material is the material on homeostatic property cluster definitions described in the Introduction and developed in the sections entitled "Metaphors, property homeostasis, and deference to nature" and "Homeostasis, reference, and precision." This material is also developed in Boyd (1988, 1989, 1991).

I have not made any attempt to survey in the present version the extensive literature on naturalistic conceptions of knowledge and of reference that has appeared since the first version was published, nor have I surveyed the equally extensive literature on the relation between naturalistic conceptions of reference and issues in the philosophy of mind and the philosophy of psychology. The reader interested in recent developments might start with Burge (1986), Devitt (1981), Dretske (1981), Fodor (1981), Goldman (1986), and Stalnaker (1984).

Since Professor Kuhn has not rewritten his comments in light of the new material in the present version I want to say something about the relation of the new material to my disagreement with him over the relative merits of realist and social constructivist interpretations of scientific knowledge. I take the development of the homeostatic property cluster theory of (some) natural kind definitions to be important to the articulation of a naturalistic conception of scientific knowledge and of the semantics of scientific language on which the defense of realism ultimately depends. In that sense *only* I believe that it contributes to the defense of realism against social constructivism. The acknowledgment and articulation of a version of the homeostatic property cluster account of certain kind definitions is plainly compatible with social constructivism. After all, social constructivists are no more Humeans about causal structures than are realists, and they can certainly portray scientists as defining some terms in terms of causally determined property correlations in the world(s) they study. Nothing in the present essay is designed to show that constructivist accounts of how such accommodations to causal structure are secured in scientific research must be inferior to realist versions. I discuss the relative merits of realist and constructivist accounts of such accommodation in Boyd (1990a, 1990b, 1991) and especially in Boyd (1992).

[1] Here, and in other entries on this list, I have abused the use–mention distinction. The reader will have no difficulty in providing (somewhat tedious) but correct reformulations of these points.

[2] It is important to understand in just what respect natural kinds are "objective" or "independently existing." According to the account offered here, natural kinds are discipline- or interest-relative. That is, the "naturalness" of a natural kind consists in the fact that its members have relevantly similar causal powers (or causal histories, etc.). Relevance of similarity is assessed with respect to the sorts of everyday reports, inductive generalizations, or theory constructions that are required for the particular practical or theoretical projects that the relevant

linguistic community undertakes. Thus "jade" denotes a commercial and "gemological" natural kind, even though for purposes of geology, jadite and nephrite are quite distinct kinds. Indeed, the notion of a natural kind can be fully explicated in terms of the notion of linguistic accommodation in the setting of particular practical or theoretical projects.

This project relativity of natural kinds represents the *only* grain of truth in Locke's claim that, although nature makes things similar, men rank them into kinds (Locke, Book III, Chap. iv, Sections 35–8). In no other respect are kinds un-"objective." The causal structures to which our language is accommodated exist quite independently of our conceptual schemes or theory construction. We do not decide by convention where the boundaries of natural kinds lie. Neither do we, in any important sense, "construct" the world when we adopt linguistic or theoretical frameworks. Instead *we* accommodate *our* language to the structure of a theory-independent world (contrast Kuhn, 1970a; Putnam, 1977; for further discussion see Boyd, 1990b, 1991, 1992).

22

Metaphor in science

THOMAS S. KUHN

If I had been preparing the main paper on the role of metaphor in science, my point of departure would have been precisely the works chosen by Boyd: Max Black's well-known paper on metaphor (Black, 1962b), together with recent essays by Kripke and Putnam on the causal theory of reference (Kripke, 1972; Putnam, 1975a, 1975b). My reasons for those choices would, furthermore, have been very nearly the same as his, for we share numerous concerns and convictions. But, as I moved away from the starting point that body of literature provides, I would quite early have turned in a direction different from Boyd's, following a path that would have brought me quickly to a central metaphorlike process in science, one which he passes by. That path I shall have to sketch, if sense is to be made of my reactions to Boyd's proposals, and my remarks will therefore take the form of an excessively condensed epitome of parts of a position of my own, comments on Boyd's paper emerging along the way. That format seems all the more essential inasmuch as detailed analysis of individual points presented by Boyd is not likely to make sense to an audience largely ignorant of the causal theory of reference.

Boyd begins by accepting Black's "interaction" view of metaphor. However metaphor functions, it neither presupposes nor supplies a list of the respects in which the subjects juxtaposed by metaphor are similar. On the contrary, as both Black and Boyd suggest, it is sometimes (perhaps always) revealing to view metaphor as creating or calling forth the similarities upon which its function depends. With that position I very much agree and, lacking time, I shall supply no arguments for it. In addition, and presently more significant, I agree entirely with Boyd's assertion that the open-