Truth, Epistemic Ideals and the Psychology of Categorization[1]

Robert N. McCauley

Emory University

Many philosophers, and particularly those of a Kantian stripe, have suspected that claims about the ways the world and its joints are (in contrast to claims about the joints in our models of the world and the way they are) make little sense. Recently, similar views have gained prominence in the philosophy of science and in cognitive psychology. The demise of the strong distinction between theoretical and observational terms over the past two decades is just one, among many, developments in the philosophy of science which is of a piece with such suspicions. Meanwhile, a great deal of recent theoretical work on the psychology of categorization also affirms the fundamentality of our cognitive constructs in structuring human experience.

In cognitive linguistics (Lakoff (forthcoming)) and in both the information processing approach (Murphy and Medin 1985) and the Gibsonian ecological approach (Neisser 1985 and (in press)) in cognitive psychology researchers have (independently) offered extensive proposals about the sorts of cognitive constructs required to account for a host of recent experimental findings concerning categorization. These researchers have undertaken these theoretical excursions because the classical view of categorization (which holds that empirical categories are defined by a set of necessary and sufficient conditions) and the empiricist learning theory that regularly accompanies it (which holds (1) that these necessary and sufficient conditions are constructed from a set of cognitively simple properties and (2) that basic set theory contains all of the necessary resources for modeling the psychological processes involved in categorization) have generally proven incapable of accomodating a number of recent experimental findings and particularly those concerning the graded structure of empirical categories and the cognitive precedence of the so-called basic level categories.

These new approaches to the psychology of categorization overlap considerably, despite the fact that they describe our cognitive constructs in somewhat different terms. Lakoff talks about idealized cognitive models while Murphy and Medin use the term "theory". (Like Neisser, I will employ both notions depending upon the context, assuming that theories are simply the more elaborate and articulated of our idealized cognitive models.) The crucial point, though, is that all of

---

PSA 1986, Volume 1, pp. 198-207

these researchers insist on the centrality of such larger cognitive constructs in their accounts of the experimental results in question. What I will call 'the idealized cognitive model/theory approach'[2] to categorization emphasizes the role that our cognitive models play in perception and categorization and denies the possibility of cognitively unmediated access to external (or, for that matter, internal) reality (Churchland 1984).

The epistemological implications are familiar. This view seems to suggest (following Quine 1951) that a constraint on the adequacy of any semantic analysis is that it include some account of the role of the larger conceptual scheme in question taken as a whole. The unit of cognitive significance is the whole conceptual scheme. Unraveling the resulting tangle concerning the point of entry in semantics has preoccupied much recent epistemology.

Nonetheless, this general approach to the psychology of categorization need not jeopardize accounts of scientific rationality that argue for scientific progress in terms of improvements in our theories' descriptions of the world. That we cannot get beyond our accounts of the world to the world-in-itself perceptually need not mean that we cannot have as good grounds as we could ever expect to have for holding that some of our accounts of the world are true or are, at least, closer to the truth than others.

To offer a detailed discussion about why we come to hold some systems of claims in such high esteem is not my goal. That would involve a discussion of issues no less complex than the sum of the considerations in light of which communities (and scientific communities especially) come to prefer the theories that they do. Rather, my goal is to defend the epistemic importance of a notion of truth in terms of a claim's correspondence with the world, the inescapability of our cognitive constructs' influences notwithstanding.

Our theories and models employ categories and infuse them with meaning in their systems of claims about the world. What we make of these categories depends upon whether or not we take the claims of those theories and models to be true. But, of course, our views about the truth of claims often change, whereas those claims' actual truth or lack of it (particularly in the case of universally quantified, law-like claims) should not.

Ultimately, this is probably the most important consideration motivating investigations of truth. Indeed, both the vitality and the adequacy of larger epistemological projects turn, in part, on their accounts of (the apparent inevitability of) this partial incongruence between our judgments and the truth. Putnam explicates this incomplete fit in terms of the distance between our actual knowledge seeking activities and our epistemic ideals. Those ideals on his view are pivotal in a satisfactory account of truth: "If the notion of comparing our system of beliefs with unconceptualized reality to see if they match makes no sense, then the claim that science seeks to discover the truth can mean no more than science seeks to discover a world picture which, in the ideal limit, satisfies certain criteria of rational acceptability. . . . truth itself gets its life from our criteria of rational acceptability. . . ." (1981, p. 130). Truth cannot be explicated in

terms of our contemporary standards of rational acceptability, because these too often change with advances in our knowledge (Putnam 1981, p. 55). Epistemology no less than science uses idealizations, and such a strategy is no more nor less problematic than those science employs. Because such epistemically ideal conditions (Putnam 1983, pp. 84-85) never completely obtain, we too often prove to be the fallible inquirers that we too often forget we are. Nonetheless, we can approximate such idealized conditions for some statements and especially for those that deal with states of affairs which seem transparent to us perceptually.

One of the most central considerations in assessing the rational acceptability of claims concerns their relation to what we perceive and (derivatively) the conditions under which that perception occurs--hence the preoccupation of so much recent epistemology with observation sentences. Ideally, perception is nonproblematic when the world reliably proves to be just as it appears. Crucially, not only do our cognitive constructs not preclude this sort of perceptual transparency, they are essential to it (Heil 1983). It is only where we have such constructs (in a very broad sense of that term which includes our innate equipment--see McCauley (in press)) that the world's contents can become transparent to our perception. Precisely because of our theories, we know both what to look for and what is similar to what in the pertinent domains. It is in such situations (i.e., those where we are not only clear about what we see, but also where interference is at a minimum, where we are not under the influence of mind altering drugs, where we are not observing a magic show, etc.) that we approximate, in at least this one crucial respect, the sort of ideal conditions with which Putnam would presumably be concerned.

Reflection on both scientific change (e.g., Kuhn 1970) and the psychology of categorization discloses misleading simplicities in this account of what I have been calling perceptual transparency. It is a consequence of the idealized cognitive model/theory approach to categorization that what we take to be the truth of claims about what virtually anyone can see to be the case actually depends upon how well those claims both cohere with the theoretical edifice that constitutes our background knowledge and (in part therefore) correspond to what we perceive. This is akin to Putnam's claim that " . . . statements are not 'made true' . . . by mind-independent states of affairs, but by states of affairs as perceived and conceptualized. . . ." (1983, pp. 83-84). Although in this passage Putnam describes this view as "non-realist" and he rejects an account of truth as correspondence, these moves are rather misleading on two counts. First of all, Putnam usually refers to the resulting position as "internal realism". Furthermore, his position does not preclude claims corresponding with the world, but only that they could ever delineate unconceptualized reality, i.e., the world-in-itself.

Admittedly, once it is unhooked from what Putnam calls "metaphysical realism" (i.e., the view which holds that we can have access to unconceptualized reality), what survives of a correspondence notion of truth inevitably must conceal a good deal of complexity. (This is what has driven its pragmatist, coherentist, and experientialist detractors.) A straightforward theory of truth as correspondence no more exhausts our use or understanding of the concept 'true' than our theories of marriage or kinship exhaust our use or understanding of the concepts 'bachelor'

or 'mother' (Lakoff (forthcoming)). In most areas (including the philosophical and the scientific) we employ any of a number of related, but not fully consistent, models depending upon the context and our purposes--permitting multiple, partially incompatible, but rationally acceptable, models--specific appeal to which depends upon the problems at hand. Putnam holds (1981, pp. 72-73) that such a pluralism constitutes "trouble" for truth as correspondence not because "correspondences between words or concepts and other entities don't exist, but . . . [rather because] too many correspondences exist."

It will help to distinguish[3] between a criterion of truth and a theory of truth. Criteria of truth concern the standards we, in fact, employ in a given situation to judge the truth of a claim. This is in contrast to the larger questions which theories of truth address which concern, among other things, the epistemological grounds for both why and when we employ whatever criteria of truth we do. Putnam construes truth in terms of our epistemic ideals. However, because such ideals are probably not much more fixed than our actual standards and practices, for an internal realist in particular, nothing precludes the possibility of a theory of truth countenancing multiple criteria of truth, depending upon how closely the situations in question approximate the conditions those ideals specify. (Of course, this view follows on the idealized cognitive model/theory account of categorization and also accords with the experimental findings concerning the graded structure of categories.)

Arguably, Putnam exaggerates the troubles that internal realism and the pluralism of models it presupposes pose for the contribution of a notion of correspondence in a theory of truth. In the space that remains I will examine considerations which suggest that in an adequate theory of truth in terms of our ideals of rational acceptability the centrality of a criterion of truth in terms of correspondence is psychologically inevitable, rationally indispensable, and (therefore) epistemologically fundamental. To the extent that the following considerations condition our epistemic ideals, a theory of truth, and especially one formulated in terms of such ideals, ignores them only at its peril.

An idealized account of perceptual transparency informs our most fundamental criteria both for assessing the rational acceptability of empirical claims and (therefore, on Putnam's account) for understanding what the truth of such claims amounts to. Seeing is believing. Ideally, those aspects of the world that are perceptually transparent are the ones we know best. It is those claims that describe states of affairs that "virtually anyone can see to be the case" about whose truth we are most confident--regardless of how sophisticated the theories which influence that seeing happen to be. This account of perceptual knowledge requires no more complex a notion of truth than one that holds that a claim must faithfully describe the way the part of the world in question is, which, of course on this idealized account, is just as it appears.

These idealized accounts of perception and truth constitute the foundation of our view of ourselves as capable of acquiring empirical knowledge. Their simplicity and coherence supply these idealized models with a psychological prominence which has insured that at the core of

virtually every alternative theory of truth has been at least a touch of correspondence.

Bare psychological prominence, however, does not guarantee bare epistemological import. A strategy for justifying these ideals depends minimally upon demonstrating that certain of our empirical concepts and at least some of the simple claims in which they figure pertain to states of affairs which are simultaneously transparent to perception and robust across a wide range of alternative conceptual schemes. The obvious question, once we acknowledge the inevitable influence of our cognitive constructs, is whether among our systems of empirical categories any such overriding stabilities hold across all or nearly all of the (rationally entertainable) alternatives.

That human beings and their perceptual systems are basically built alike offers some hope on this count. So too do the various experimental findings that indicate that at least some components of our systems of empirical categories do, in fact, persist in the face of theory change. No doubt, this indicates that reasonably detailed physiological biases constrain the way our perceptual systems cut up the world and that these biases are, in Pylyshyn's (1980) terms, cognitively inpenetrable.

The findings concerning basic level categories are relevant here (Rosch et al. 1976 and Mervis and Rosch 1981). These categories are cognitively salient on any number of independent empirical criteria. (In the hierarchy of toy poodle, poodle, dog, mammal, animal, organism--dog is the basic level category.) In any class inclusion hierarchy of empirical concepts we can identify a basic level. The term for the basic level is most likely to be learned first, to be orthographically and phonologically least complex, and to be recognized and remembered most easily (controlling for such variables as word length and word frequency). The basic level also represents a unique peak in the increase of information when moving from one level to another in a class inclusion hierarchy. A large number of these categories' properties distinguish them from their immediate superordinates but very few distinguish them from any of their subordinates. Finally and, surely most importantly, the basic level is the most superordinate at which subjects can still readily provide either a mental or physical image of the relevant categories' members.

Not only do class inclusion hierarchies of empirical concepts always have some basic level, even more importantly, for middle-sized, terrestrial objects, i.e., what comprises in large part the world to which we have adapted, most basic level distinctions seem to be quite robust across a wide variety of languages, subjects, settings, and situations (Lakoff (forthcoming)). Although the situation is less clear-cut (Dougherty 1978) in special cases of personal or cultural expertise (where the relevant basic level distinctions sometimes seem to shift on certain criteria to a more subordinate level), there is no evidence (contra Murphy and Medin 1985, p. 305) that subjects' basic level distinctions change on any of the perceptual criteria.

Basic level categories seem to define some of the world's major joints. That these distinctions are perceptually transparent and that they are the most superordinate categories for which we can identify

typical patterns of interaction with the objects that they pick out (Neisser 1985 and Lakoff (forthcoming)) are evidence of well entrenched models at work. That so many persist in the face of considerable contextual, theoretical, and cultural diversity suggests that a relatively stable and extensive set of expectations informs (without conscious reflection) many aspects of human perception and categorization. (A view which so emphasizes the role of our cognitive constructs in perception and categorization does not preclude perception of either the novel or the unexpected. Still, such anomalies need not, indeed cannot, violate the natural constraints which operate in human perception. It only follows that we will not automatically know specifically either what to make of them or do with them. They will not be transparent, because they violate the system of expectations of the theory with which we are presently operating. No theory of cognitive functioning should rule out the possibility of such anomalies. Because of such experiences we choose either to employ alternative theories at our disposal or to consider constructing wholly new ones.)

Of course, theoretical advances in science could modify or even overturn various basic level distinctions piecemeal. They probably have, on occasion (Gould 1983). Nonetheless, in light of what are probably physiological limitations and constraints on us as knowers, certain empirical distinctions seem to have proven so central to our negotiations with the world that they appear virtually unrevisable. Theoretical change need not engender uniform upheaval throughout our entire system of knowledge. Some distinctions may be so integral to making our way in the world that they will not, in fact, ever be surrendered. Empirical categories which prove relatively immune to both theoretical diversity and theoretical change would constitute a system of generally fixed points in our knowledge scheme.

These claims should be read carefully. It is not that these basic level distinctions constitute either given or ultimate standards of reliability or truth--in the sense Popper correctly criticizes (1972, pp. 63-72). Justifying a central role for a correspondence criterion of truth in an idealized theory of truth by way of our basic level distinctions requires neither that such categories offer clues about the character of the world-in-itself nor that any particular category among this set be theoretically insuperable. In one sense we do employ these categories dogmatically, but the fact that "this more dogmatic part of common sense" (and its systemic stability) "is always a good starting-point" (Popper 1972, p. 69, emphasis mine) is not only a fact for which our epistemological positions should account but also one which has important implications for a theory of truth.

Putnam cites the fallibility of our physiological biases and the limitations of our experience as the bases for claiming that " . . . even principles we regard as 'a priori' . . . will from time to time turn out to need revision in the light of unexpected experiences or unanticipated theoretical innovations." (1981, p. 83). However, he also rejects the possibility of unlimited conceptual revision, because it would undermine even the possibility of a coherent notion of rationality. Without some long-term, readily familiar stability in our ever emerging theoretical edifice, and, consequently, in the world that we perceive, it would be quite difficult to justify employing either the idealized epistemological models I have been discussing or

(correspondingly) the notion that some aspects of the world are transparent perceptually. It is revealing, though, that it would be comparably difficult to explain our having either empirical knowledge or rational discussion (at least about empirical matters).

Even disputants must agree about what certain portions of the world are like. Where alternative models for a particular domain overlap delineates these points of agreement. I have been suggesting that evidence from experimental psychology indicates that the stability even among our system of more clearly empirical concepts exceeds what we would expect, if all of our knowledge was really up for grabs. Like Davidson's, this approach "is not designed to eliminate disagreement . . . [rather] its purpose is to make meaningful disagreement possible, and this depends entirely on a foundation--some foundation--in agreement." (1974, pp. 196-197). Those "dogmas" or points of agreement which no conceptual scheme seems able to avoid are either grand mistakes, partial constraints on the very possibility of rational exchange, or, conceivably, both. Surely, other things being equal, the obligation to preserve these particularly robust empirical distinctions partially constitutes our ideal standard of rational acceptability in our discussions of empirical matters. Our basic level distinctions support an appropriately restricted notion of truth as correspondence precisely because "truth itself gets its life from our criteria of rational acceptability" (in the ideal limit).

Things could not have come out any other way, given the sorts of organisms we are (viz., ones whose cognitive constructs (in the broad sense) substantially constrain what we perceive) and given the way that we must manage in the world (viz., where that perceptual input is our sole input). It is difficult to conceive, short of Cartesian demons, how the world that we can know could ever be so independent of what we think (and especially of what we think about our perceptual input) that even many of our basic level distinctions could be misleading simultaneously. Such basic categories in our system of knowledge cannot even be mostly in a flux at once; otherwise we would lose all sense of the world which is their object.

I concur with Putnam that we can never be positive about those distinctions which would constrain any of our revisions. I am, nonetheless, suggesting that we have some reason to hold that for any conceptual scheme whose claims will overwhelmingly fall among the "rationally acceptable" the vast majority of our basic level distinctions are inescapable.

The force of these claims is not merely psychological. The epistemological prominence of truth as correspondence is supported, in part, because upon continuing reflection the members of basic level categories, in fact, continue to prove to be, for the most part, what they seem to be. The staying power of our basic level categories provides tremendous impetus for realism about them, at least. That realism undergirds our idealized model of perception as transparent, which, in turn, undergirds the centrality of correspondence in a theory of truth. (Often, we can simply see that some claims are true.) The closer our understanding of a claim's truth approximates the ideal of perceptual transparency (and the satisfaction of a correspondence criterion of truth) the greater our confidence in its truth is. Of

course, the centrality of the theories that address the phenomenon in question (to our overall system of empirical knowledge) substantially determines our sense of the proximity of a claim's truth conditions to the ideal of a correspondence criterion of truth.

I am suggesting, then, that these idealizations also play a central role in our conception of epistemically ideal conditions. Although it does not exhaust our understanding of truth, a notion of truth as correspondence serves as a regulative ideal epistemically. We can never completely shake a view of truth as correspondence. For an internal realist in particular this fact should have epistemological import. For it is precisely here both where psychological and epistemological matters converge and, consequently, where we can lay some claim to having met Quine's recommendation for a naturalized epistemology--without committing any naturalistic fallacy, Putnam's (1983) claims to the contrary notwithstanding.

The notion that we cannot escape our theories leaves us, perhaps, with a rather unimpressive account of "the world" as simply those objects (and their relations) in our theoretical edifice which, for the moment, we leave unquestioned. However, the fact that our theoretical advances have not rendered our basic level distinctions obsolete, certainly not for practical purposes nor for most of our theoretical purposes, offers some reassurance. We have a theoretically robust set of provisionally fixed points which constitute the rough outlines of a world to which our new theories' claims should either correspond or explain away.

Notes

[1]I wish to express my gratitude to Larry Barsalou, William Bechtel, John Heil, George Lakoff, Douglas Medin, Ulric Neisser, and Robert Richardson for their helpful comments. Positions advanced in this paper are closely related to those in the last section of McCauley (1985).

[2]I regret this somewhat concocted name, but it does have the virtues of remaining both faithful to all of the texts in question and transparent with respect to the positions' most fundamental substantive commitments.

[3]I am grateful for conversations with Charles Nussbaum on this topic.

References

Barsalou, L. (1982). "Context-independent and Context-dependent Information in Concepts." Memory and Cognition 10: 82-93.

Churchland, P. (1984). Matter and Consciousness. Cambridge: The MIT Press.

Davidson, D. (1974). "On the Very Idea of a Conceptual Scheme." Proceedings of the American Philosophical Association XLVII: 5-20. (As reprinted in Inquiries into Truth and Interpretation. Oxford: Oxford University Press, 1984. Pages 183-198.)

Dougherty. J. (1978). "Salience and Relativity in Classification." American Ethnologist 5: 66-80.

Gould, S. (1983). "What, If Anything, Is a Zebra?" In Hens' Teeth and Horses' Toes. New York: Norton. Pages 355-365.

Heil, J. (1983). Cognition and Perception. Berkeley: University of California Press.

Kuhn, T. (1970). The Structure of Scientific Revolutions. 2nd ed. Chicago: University of Chicago Press.

Lakoff, G. (forthcoming). Women, Fire, and Dangerous Things: What Categories Tell Us About the Nature of Thought. Chicago: University of Chicago Press.

McCauley, R. (1985). Concepts, Theories, and Truth. Emory Cognition Project Report #8.

----------. (in press). "The Role of Theories in a Theory of Concepts." In Neisser (in press).

Mervis, C. and Rosch, E. (1981). "Categorization of Natural Objects." Annual Review of Psychology 32: 89-115.

Murphy, G. and Medin, D. (1985). "The Role of Theories in Conceptual Coherence." Psychological Review 92: 289-316.

Neisser, U. (1985). From direct perception to conceptual structure. Emory Cognition Project Report #7.

----------. (ed.). (in press). Concepts Reconsidered: The Ecological and Intellectual Bases of Categorization. New York: Cambridge University Press.

Popper, K. (1972). Objective Knowledge. Oxford: Oxford University Press.

Putnam, H. (1981). Reason, Truth and History. New York: Cambridge University Press.

----------. (1983). Realism and Reason. (Philosophical Papers, Volume 3.) New York: Cambridge University Press.

Pylyshyn, Z. (1980). "Computation and Cognition: Issues in the Foundations of Cognitive Science." The Behavioral and Brain Sciences 3: 111-132.

Quine, W.V.O. (1951). "Two Dogmas of Empiricism." The Philosophical Review 60: 20-43. (As reprinted in From a Logical Point of View. New York: Harper & Row, 1963. Pages 20-46.)

Rosch, E.; Mervis, C.; Gray, W.; Johnson, D.; and Boyes-Braem, P. (1976). "Basic Objects in Natural Categories." Cognitive Psychology 8: 382-439.