SEMANTICS AND PSYCHOLOGICAL PROTOTYPES

BY

BERNARD W. KOBES

Young woman in museum: What's that bird?
W.P. Ker: It's a guillemot.
Young woman: That's not my idea of a guillemot.
W.P. Ker: It's *God's* idea of a guillemot.
<div align="right">Geoffrey Madan, Notebooks</div>

The past two decades have seen the development of "anti-individualistic" accounts of intentional content which seek to explain the content of thought and speech partly by reference to the nature of the thinker/speaker's physical and social environment. Hilary Putnam, Tyler Burge, and others have argued that what a thinker or speaker thinks or means is "fixed" partly by states of the world external to the individual. The point is not the obvious one that environmental stimuli cause us to have certain concepts and not others. Rather, it is that by conceiving a thinker's environment as varying we can vary the intentional type-identity of his concept while his inner states and dispositions (non-intentionally described) remain constant. The natures of things which the thinker and members of his community actually perceive and manipulate help fix the type-identity of his concept.[1]

  In the same period psychologists have devoted a great deal of attention to the representation of everyday object categories. Investigations by Eleanor Rosch, Amos Tversky, Douglas Medin, and others, indicate the importance of variation in "typicality" in our representations of object categories, and in our categorization procedures.[2] An instance or subcategory is typical with respect to a containing category to the degree that it is a "good example" of the containing category. A variety of experimental results suggest that it is a psychologically important fact

about our representations of object categories, and about our categorization procedures, that (for example) *sparrow* is a more typical subcategory of *bird* than is *ostrich*.

Philosophers have exerted remarkably little effort toward articulating a theoretical framework in which these two bodies of research can be seen in relation to one another.[3] Such a framework seems desirable for two reasons. First, the theories share some emphases: for example, both claim a large role for the perception and manipulation of actual objects in shaping the identities of our everyday concepts, and a small role for the representation of necessary and sufficient conditions. Second, the psychological literature is full of philosophical groping about in foundational matters. Psychological theorists are concerned, to varying degrees of explicitness, with such questions as: What sort of thing is a concept? And what is the relation between descriptive statements giving the range of things to which a subject is disposed to apply his concept, and normative statements about the range of things to which a subject's concept correctly applies?

My goals in this paper are two. The first requires a bit of stage-setting from the psychological literature. Some prototype psychologists have advanced confused accounts of concept identity, accounts which have been justly criticized in the experimental and theoretical literature. I shall argue, in section I, that their error lies not in taking prototype effects seriously in a theory of concepts, but in a confusion as to the appropriate aims of psychological theory. By assuming, as anti-individualistic semantic theories do, that we have epistemic access to "objective" categories (roughly, categories whose integrity is independent of the subject's beliefs, dispositions, and categorization procedures) we can specify more limited but still interesting explanatory aims for prototype theory.

My goal in sections II and III is to explain and argue for a sense in which semantics and psychological prototype theory are mutually dependent.

I.   Prototype Effects and Objective Categories

The key experimental finding of the prototype literature is that instances of everyday object categories have degrees of typicality or goodness-of-example that are fairly constant across subjects within a culture, and that a number of distinct experimental tests yield similar typicality rankings.[4] Experimentally derived typicality ratings are highly correlated; i.e., subjects tend to agree about the relative typicalities of instances of a category.[5] Knowledge of the more typical instances of a category is acquired sooner in a child's conceptual development than knowledge of less typical instances.[6] Children acquire many everyday concepts by generalizing from central examples that are given as instances; frequently

they over- or under-generalize. When subjects are asked to list instances of a category, the probability that an instance will be listed, and the order in which instances are listed, correlate with typicality.[7] Finally, there is a striking psychophysical correlation: typical instances are categorized more quickly than less typical instances.[8] For example, in a timed categorization task, it takes subjects longer to confirm that a spider is an animal than to confirm that a dog is an animal; and it takes them longer still to confirm that a protozoan is an animal.

Rated typicality within a containing category correlates with measures of "family resemblance" based on the number of psychologically salient features the relevant instance shares with other members of the category. For example, define F(a,C), a (crude) measure of the family resemblance of instance a with respect to the containing category C, as follows:

$$F\,(a,C) = \sum_{b \in C} N\,(a,b),$$

Here C is identified with the set of its instances (including a). N(a,b) is the number of psychologically salient features shared by a and b. Thus F(a,C) is the number of salient features that a shares with other instances of C. Equivalently, if each salient feature of a is weighted according to the number of instances of the containing category that have the feature, then F(a,C) is the sum of the weights of a's salient features.[9]

Experimental work in the prototype tradition has also uncovered a psychologically important "vertical" structure in taxonomies of categories. There is a basic level of abstraction that picks out the most inclusive categories whose instances have roughly the same shape. The basic level categories are neither the most inclusive nor the most specific. Categories more abstract than the basic level are called "superordinate"; those more specific than the basic level are called "subordinate". For example, hammer is basic, tool is superordinate, and claw hammer is subordinate. Besides being the most abstract categories representable by a summary image, basic level concepts are acquired first in conceptual development. The basic level is where the child "enters" the taxonomy. Moreover, there is evidence that objects are visually recognized first as members of their basic category. For example, subjects see an object first as a hammer, and only later as a claw hammer, or as a tool.[10]

Prototype theorists attempt to account for the experimental effects by postulating structures for concept representation and procedures for arriving at categorization judgments. Competing theories emphasize binary features, geometrically continuous dimensions, particular exemplars, and idealized dynamic models respectively. I shall not be concerned here with these theoretical differences. The extent to which the theories compete with each other is sometimes overestimated; they frequently

concern distinct aspects of representation and categorization. What the theories have in common is that they postulate representations and procedures that induce psychological centrality and distance effects among instances of everyday concepts.

The prototype phenomena have suggested to some theorists certain semantic, even metaphysical, theses. The experimental effects have been taken to reveal the structure of the range of things to which a subject's concept applies, in a sense of 'applies' that should interest semantics. The range of *correct* application of a subject's concept is structured by a gradient of psychological similarity to "central" subcategories or instances. Consider, for example, the following passage from an influential paper of the mid-1970's:

There is increasing evidence that membership in the semantic categories referred to by the words of natural languages is not an all-or-none phenomenon. Contrary to the assumption that categories are necessarily logical, bounded entities, . . . many natural categories are continuous and possess an internal structure in which members are ordered according to the degree to which they are judged good examples (typical) of the category.[11]

On this "semantically loaded" view (as I shall call it), traditional philosophers, logicians, and psychologists have forced concepts into a Procrustean bed in supposing that there must be severally necessary and jointly sufficient conditions corresponding to each (non-primitive) concept, and concomitantly that membership in the extension of a concept is binary. That is to say, any given item is either in the concept's extension, or not, with no grades of category membership in between. But the concepts studied in prototype theory seldom have associated necessary and sufficient conditions. Such conditions do not figure in prototype accounts of concept representation, nor in the corresponding accounts of categorization. Moreover, on the semantically loaded view, experimental results such as the family resemblance work show that categories themselves, "out there" in the world, have smoothly graded rather than binary structures. Concept extensions are therefore fuzzy sets in the sense of Zadeh's "fuzzy set theory", which assigns to each element a numerical value in the interval [0,1], where 0 represents the least possible, and 1 the greatest possible, degree of membership in the set.[12]

The semantically loaded view conflates typicality with category membership. For the commentary makes an unacknowledged and unsubstantiated leap from smooth gradations of typicality to smooth gradations of category membership. This point is made by Georges Rey, who notes that we must distinguish a man's being bald to degree 0.3 from a subject's being confident to degree 0.3 that the man is bald. Since typicality can vary due to epistemological factors with no direct connection to semantic

("metaphysical") factors, we cannot infer gradations of category membership from gradations of typicality.[13]

Granted, everyday object concepts have unclear boundaries, and given almost any artificial sharpening of a concept, one can imagine further unclear cases. Unclear boundaries sometimes play a role in determining degree of typicality, as in the cases of *euglena* with respect to the containing category *animal*, and *carpet* with respect to *furniture*. However, the role is minor; other determinants of typicality will overwhelm unclear boundaries. Certainly unclear boundaries are not the best explanation for the sorts of typicality rankings derived from the experimental paradigms of the prototype literature.

Daniel Osherson and Edward Smith argue against prototype theory from the observation that the typicality of an object with respect to a conjunctively defined category may exceed the typicality of the object with respect to either of the conjuncts.[14] The example they use is the conjunctive concept *striped apple*. A particular striped apple may be a "good example" of the concept *striped apple*, but atypical of the concept *striped*, and atypical of the concept *apple*. Its typicality with respect to the conjunctive concept therefore exceeds its typicality with respect to either conjunct. But according to fuzzy set theory, the degree of membership in an intersection of sets is the minimum of the degrees of membership in the component sets. If concept extensions are fixed by psychological typicality, and if fuzzy set theory is used to fix the extensions of logically complex concepts, we have a contradiction.[15]

Critics of prototype theory have often taken the semantically loaded gloss to be the heart of prototype theory. Consider, for example, Osherson and Smith's formal characterization of prototype theory. Prototype theory, on their account, identifies concepts with quadruples $<A, d, p, c>$, where A is a set of "readily envisionable" objects, d is a "psychological distance" metric mapping $A \times A$ into the non-negative reals, $p \in A$ is the concept's prototype, and c is the concept's characteristic function, mapping A into [0, 1] such that

$$[d(x,p) \leq d(y,p)] \rightarrow [c(y) \leq c(x)].$$

This condition, according to Osherson and Smith, requires that "the closer an object is to its prototype, the more characteristic it is of the concept".[16] The intent is to *explain* the characteristic function c, which gives the extension of the concept (a semantic notion), in terms of prior notions of psychological distance d, and psychological centrality. This gives rise to the problem of conjunctive concepts, which Osherson and Smith then wrongly take to be a difficulty with prototype theory quite generally.

It is possible and desirable, I shall argue, to develop an interesting prototype theory without commitment to the semantically loaded gloss.

Note, first of all, that one can combine relatively binary categories with prototype theory by treating category boundaries as falling at a threshold distance from the prototype for the category. Several categories may be defined simultaneously as a "contrast set", such that an item falls into the extension of one of the categories only if it is closer to the prototype for that category than it is to the prototype for any other category in the contrast set.[17] This sort of account has plausibility for some terms—for example, color terms.

I want to focus, however, not on the question of whether our everyday concepts are in general sharply binary; even in light of the work of Putnam and Burge, it is quite clear that they are not. I shall focus instead on the deeper assumption of the semantically loaded gloss, namely, that category boundaries, and hence a notion of application that should interest semantics, can be *explained* in terms of prior notions of psychological centrality and psychological distance. It is evidently this assumption that motivates the move from smoothly graded typicality to smoothly graded category membership.

On Putnam's account of the semantics of natural kind terms, the way the world actually is helps fix the extension of a natural kind term, in such a way that the extension can be given a conditional ostensive characterization: if *this* sample is in fact water, then necessarily x is water if and only if x bears the same-liquid relation to it. The point applies not only to terms, but also to concepts. The identity of an individual thinker's concept cannot be explained purely by reference to his associated beliefs, dispositions, and abilities. For it is partly a matter of some relevant shared nature of actual things to which the concept has been applied.[18]

On the resulting picture, simplified and extended to terms for artifactual and social kinds, an object concept is applied by members of the community, in conjunction with perceptual and behavioral interaction, to a set of things that in fact share an underlying nature. (I use the term 'nature' broadly, and for the moment somewhat vaguely, in a way that allows that artifactual and social kinds may have a nature.) Applications of a concept will be linked to each other, and to perceptually based applications in particular, by tacit intentions of the thinker. I shall use the term 'sample' to refer to the set of existing things to which the concept has actually been applied, by the thinker or other members of his cognitive community.

The concept correctly applies to all things that share the nature of the sample. Thus a creature falls into the extension of *penguin* [*vegetable*, *wedding*, . . . mutatis mutandis] if and only if it is of the same kind as, shares the nature of, the creatures to which *penguin* [*vegetable*, *wedding*, . . .] has actually been applied. (This must not be confused with the circular proposal that a creature falls into the extension of *penguin* if and only if it is capable of having *penguin* correctly applied to it.)

This sketch of extension-fixing allows for, and ignores, all the typicality variation you please in the concept's extension. The account alludes to the shared nature of the sample, known or unknown, but usually "hidden", whereas psychological typicality sits on the surface. The fact that the shared nature will usually be psychologically inaccessible ensures that the concept's extension will not be fixed by accessible typicality patterns in the sample. The extension will include equally all instances, however atypical, that share the actual nature of the sample. Thus a thinker's concept *bird* will include in its extension penguins, and exclude bats, even if the thinker's concept *bat* is psychologically closer than his concept *penguin* to his prototype for *bird* (even if, indeed, he judges that penguins are not birds, and that bats are).

That the notion of "shared nature" makes no invidious discriminations based on typicality is reflected in the practice of (natural or social) scientists, whose broad purpose it is to uncover and articulate the shared nature of the kind. I assume that our ordinary thought and discourse about the world is broadly continuous with scientific thought and discourse. There is no evidence that science treats atypical instances as belonging to the extension to a lesser degree. In biology, for example, penguins, pelicans, ostriches, hummingbirds, and condors are simply birds; and cowgirls are simply females. However, the existence of the shared nature, and its role in fixing intentional content, do not depend on the existence of an articulated scientific theory.

Whether something shares the nature of the sample will itself have no clear answer in some cases. Although cowgirls are simply females, in certain cases of androgen insensitivity (in which a genetic male is phenotypically partly female) the simple binary classification scheme breaks down. My claim is not that concept extensions are absolutely binary, but that they are binary relative to the broad gradations of psychological typicality.

This account of the relation between psychological typicality and category boundaries can be expressed more generally in terms of family resemblance. Let C be an objective category of actual objects or aggregates that share a nature. In any such category there will be typicality differences resulting from the pattern of distribution of psychologically salient features; instances of C will differ in their "family resemblance measures". Recall that we defined $F(a,C)$, our (crude) measure of the family resemblance of instance a with respect to C, as follows:

$$F(a,C) = \sum_{b \in C} N(a,b),$$

where $N(a,b)$ is the number of psychologically salient features shared by a and b. The point is that instances of C will differ in the number of

psychologically salient features that they share with other instances of C. Variation in psychological typicality therefore does not in the usual case reflect the structure of the category represented, nor does it indicate a fuzzy concept extension. It merely reflects the variation in number of shared psychologically salient features that one would expect to occur in any objective category.

It has turned out that people can and do *pick up* such variation in typicality, and that this has profound psycho-physical effects. In the course of our interaction with the objective category we acquire information about the family resemblance structure inherent in it. This information in turn guides our applications of the concept. Credit for these findings, and the empirical task of extending and deepening them, fall to prototype theory. But these findings could hardly show that the original category C, the sharply bounded category of objects and aggregates that share a nature, is literally *unthinkable*.[19]

The irony of this situation is that the family resemblance results have particularly encouraged the advocacy of fuzzy sets as concept extensions. The family resemblance results do show that typicality effects are not just accidental products of the nature of our representational system, but are rooted in the psychologically significant structure of reality. But the "semantically loaded" account of prototype theory fails to distinguish this from the claim that categories "out there in the world" are smoothly graded rather than binary. We forestall the conflation by pointing out that a psychologically efficacious prototype structure is quite consistent with the assumptions that there are objective, relatively binary categories of objects and aggregates that share a nature, and that we can represent such categories.

Psychological centering and distance effects are precisely what we should expect in virtue of such phenomena as differential clustering of psychologically salient features within objective categories to which we have epistemic access. Prototype psychology can be a viable theory only if it gives up the ambition to characterize the identities of concepts in terms of the individual subject's psychological capacities and dispositions. As a branch of individual psychology it lacks the resources to characterize from the ground up the conditions of correct application of the concepts it attributes. It has, rather, the potential to describe the psychological structures of our representations and how those structures are deployed in variously successful procedures for arriving at categorization judgments.

The phenomena of semantics and those of prototype psychology are intertwined in such complex ways that a unified theory explaining both is perhaps an unrealistic hope for the foreseeable future. Illumination is more likely to come by distinguishing semantics from prototype theory (*qua* branch of individual psychology) according to their distinct explan-

atory goals. I turn now to some remarks on this theme. My thesis is that the explanatory goals of each theory presuppose the results of the other, so as to form a (benign) explanatory circle.

II. Prototype Theory Should Presuppose Semantics

On the picture I want to put forward, our everyday concepts arose historically, and are continually maintained, in virtue of our perception and manipulation of everyday objects and events. Such interaction gives rise to and maintains semantic facts about the ranges of correct application of our concepts, and psychological facts about the structures of our representations and the procedures we employ for making categorization judgments. But neither sort of fact can be wholly explained without presupposing the other sort of fact.

Prototype theory and semantics are distinct disciplines in virtue of their distinct explanatory goals. A chief goal of prototype psychology is to describe our representations and procedures with a view to explaining our classificatory successes and failures. And a chief goal of semantics for monolexical concepts is to explain how facts about our psychologies and the structure of the world give rise to facts about the ranges of correct application of our concepts. The fulfillment of each task, I shall argue, presupposes the results of the other.

As a preliminary, note that the notion of sample, the set of things to which a concept has been applied, must be distinguished from the notion of prototype. A concept's extension need not be "fixed" especially by the shared nature of the concept's most typical instances. The sample will include some highly atypical instances: we *do* apply the concept *bird* to perceived penguins, ostriches, hummingbirds, and condors. And the relation "same kind" or "shared nature" can extend from the atypical to the typical. Putnam's "ostensive definition" still works if the ostended sample is highly atypical. Suppose I ostend a young Bengal tiger of unusual coloration, a clumsy hunter, living on the fringes of the Bengal's natural habitat. It will still be true that something is a Bengal tiger if and only if it bears the relation "same subspecies" to *that* creature. It is an important psychological generalization that for teaching purposes we tend to ostend psychologically typical samples; this practice makes for efficient and informative concept learning. But if we need an ostensive expression to fix ("metaphysically") the concept and its extension, the atypical Bengal gives the right result.

Moreover, the sample and the prototype belong to different ontological categories: the sample must consist of actually existing things, but the psychological prototype is a mental representation rather than an actual animal, vegetable, or mineral. Since the prototype captures central

tendencies within a diverse group, it is frequently an average or composite, and it may be that no actual instance has all the features of the prototype. No actual dog, we may assume, matches perfectly the prototype for *dog*. The point, again, is that sample and prototype are distinct notions. But it is the same history of perceptual and manipulative interaction between thinkers and objects that marks out certain of them as belonging to a sample, and simultaneously gives rise to psychological centrality and distance relations among instances.

It is of the first importance for psychological theory to distinguish semantic properties of concepts and thoughts from causal-functional properties. The psychological literature on concepts, especially in its explicit methodological reflections, is shot through with a tacit disjunction: either (i) the notion of "correct application" of a subject's concept is irrelevant to psychological description and explanation of the subject, or (ii) "correct application" *is* relevant to psychology but can be fully explicated in terms of the individual subject's articulations and dispositions of various sorts.

I mean my slogan "prototype theory should presuppose semantics" to express an alternative to both of these. The view is that prototype psychology should attribute concepts essentially identified according to their ranges of correct application, and that a concept's range of correct application cannot be wholly explained in terms of the individual subject's articulations and dispositions; rather, it is fixed in part by the subject's natural and social environment.[20]

Prototype psychology is nevertheless, on this account, a branch of individual psychology, since it quantifies over individual psychological subjects and does not explicitly concern relations between them.[21] Moreover, the individual subject of the theory is not idealized so as to embody all the knowledge and expertise at large in his community. Like each of us, the individual subject is ignorant and error-prone. Indeed, ignorance and error under certain conditions will often be precisely the explananda of the theory.

The concepts attributed to subjects must be taken as primitive with respect to prototype theory. Our semantical viewpoint assumes that prototype theory, *qua* branch of individual psychology, has not the resources to characterize the conditions of correct application of the concept, and I have argued explicitly that the notions of psychological distance and centrality will not do the job. Yet the individual subject is committed to there being such a range of correct application, and we want to be able to explain his successes and failures by standards to which he is committed. Prototype theory therefore attributes concepts with a certain presupposed range of correct application. We employ in the language of the theory translations of the subject's general terms, for purposes of reporting in indirect discourse the subject's psychological states. The

language of the theory may have to be specially augmented with such translations.

The picture of prototype psychology put forward in Section I supports the present account of psychological explanation. The picture starts with the existence of objective, external categories. Subjects' perceptual interactions with instances simultaneously define the sample and give rise to psychological centrality and distance relations among instances. The former mechanism is in the usual case social: individuals rely on others in the community to help pick out the sample. The expertise relied on here is usually not of the most theoretical kind; it is rather an expertise of application. The relevant expertise is that of those who are in the best position to make perceptually based identifications of the kind. It would be metallurgists who are in the best position to give the application of the concept *gold*, not the most theoretical of chemists. And contextual factors will help to fix the kind. It is not necessary that experts be able to distinguish the sample from all possible mimics.

Some psychological theorists distinguish between two components of a concept: a core, and an identification procedure. On this view, possessors of a concept represent both quick classification procedures for the concept, to be used under conditions of impoverished time or data, and defining conditions (the core), the final arbiter for categorization judgments, perhaps to be used as a backup procedure for categorization under conditions of ideal access to data. The identity of the concept *woman*, for example, is jointly determined by its core (anatomical, or perhaps genetic, criteria for womanhood), and its identification procedure (dress, hair style, voice pitch, and body shape).[22]

The view is, I think, misguided. There is not particularly one procedure for identifying an instance of a concept under conditions of impoverished data; there may be many, depending on the epistemological context, or there may be none at all. Moreover, it is not plausible that a thinker can produce core definitions for each of the kind-concepts that he possesses. In a wide range of cases we are *committed* to there being an underlying nature that accounts for the salient features of the sample, without our being able say what that underlying nature is.[23]

This background commitment is often evinced by subjects' dispositions to stand corrected. We take our classificatory judgments to be at risk. Even items that seem typical instances in virtue of their perceptually accessible features may turn out to be frauds, decoys, works of *trompe l'oeil* art, Disney artifacts, or evolutionary mimics. We are educable in other ways as well: we are willing to be persuaded by experts that atypical instances are fully instances, and that our stereotypes are mistaken. Even one who stubbornly refuses to stand corrected will do so not on the ground that psychological distance relations are constitutive of the notion of correct application of his concept, but on the ground that he knows better

than others what the shared nature is. For example, someone who stubbornly resists correction of his opinion that bats are birds is more likely to seek biological evidence in his favor than to appeal to psychological distance relations.[24]

Prototype theory explains our successes and failures relative to this background commitment. The psychological theory thus takes on a normative cast. Given that our concept has a non-individualistically fixed range of correct application, how well do we fare with our collection of typicality-based heuristics? How do exemplars, similarity spaces, imagery, characteristic features, mnemonics, vivid instances, metonymy, cognitive reference points, metaphors, analogies, templates, and so on contribute to our classificatory successes? Under what conditions do we predictably fail?[25] What are the neural and evolutionary bases for our patterns of success or failure in categorization?[26] Can we characterize mathematically the value to the organism of certain kinds of heuristics under less than ideal conditions?[27]

It has been suggested that the phenomena of psychological distance are best seen as results of conceptual representations being implemented in connectionist or PDP-style neural structures.[28] Some important kinds of connectionist structures allow concept representations to be modelled as geometric spaces of instances under a psychological distance relation. The models provide a neurally plausible account of such phenomena as the faster categorization of more typical instances. Speculative as the suggestion is at present, it is worth noting that there are ways to elaborate it so as to exhibit its consistency with my account of explanation in prototype psychology.

For example: the intentional content of a connectionist structure depends not on the pattern of its internal connections or their strengths, but partly on relations the subject bears to his fellows and to environmental objects. Features of the neural implementation are exploited so as to deliver fast categorization judgments that are correct often enough to be useful. Richard Grandy has noted that for animals high on the food chain there is ecological value in a capacity to make fast judgments of opportunities to eat or be a meal, where such judgments systematically err by overestimating the (in fact small) probability of the opportunity.[29] Such a judgment can be thought of as a bet based on the best estimate the creature can arrive at under conditions of impoverished evidence and time. The creature exploits typicality structures induced by the distribution of perceptually salient features within the objective category. The connectionist story is that this is possible in virtue of the structure of the neural implementation of the concept, which codes such perceptually salient features. It is indeed plausible that resort to "inter-level" strategies (strategies that exploit features of the neural implementation) are most likely where considerations of performance speed and degraded evidence loom large.

Philosophers frequently assume an impoverished conception of the goal of psychological theory; for example, that psychology aims to explain how

sensory stimulations cause internal state transitions which in turn cause bodily motions. The upshot of our discussion is that there are interesting psychological theories in the prototype tradition which do not fit this mold. Such theories aim instead at explaining a subject's classificatory successes and failures, where success and failure are judged against public standards of correct application of the subject's concepts.

III.   Semantics Should Presuppose Prototype Theory

Where both normative theories and psychological theories exist for a range of mental or behavioral phenomena, it is frequently the case that the phenomena interact in such complex ways that neither sort of theory can be carried out in complete independence of the other. The theories are distinguished from each other not by their relative priorities in an ideally systematic order of exposition, but by their distinct explanatory goals. I conjecture that such is the case for the normative theory of rational choice and the psychology of preference and decision making, for the normative theory of epistemic rationality and the psychology of belief formation, and perhaps also for normative ethics and the psychology of moral development.

In the case of each normative theory there must be psychological facts that give point to its application to human beings. In this sense the normative theory must presuppose (parts of) the descriptive theory. So it is, I shall briefly argue, in the case of semantics, construed as a theory of how facts about our psychologies and the structure of the world give rise to facts about the ranges of correct application of our concepts.

It is a truism that meaning in some sense depends on human activity. Perception, action, and memory provide contact points with the sample, and so a full explanatory account of a concept's having its particular range of correct application will have to appeal to those psychological events. Social cooperation too must have psycholinguistic and sociolinguistic aspects; these will be presupposed in the full semantic story. But is there any specific way in which a full semantic account must presuppose the phenomena of psychological centrality and distance? Despite having argued that semantic notions cannot be reconstructed out of psychological distance relations, I believe the answer is 'yes'.

On the semantic viewpoint I have been presupposing, a subject's concept truly applies to all and only things that share the underlying nature of the sample, where the sample is picked out in virtue of causal interaction between concept users in one's community and real-world items. It is time now to confess that the notion of underlying *shared nature* is vague, and I shall not be able to make it precise enough to satisfy all reasonable scruples. I shall, however, argue for the following limited thesis: among the constraints on the relevant shared nature is that it help account for psychological centrality and distance relations in subjects'

conceptual representations. Other things being equal, the shared nature ought to provide a "deep rationale" for the typicality phenomena, on pain of there being no *point* to the normative claim that the subject's concept truly applies to all and only things of that nature.

In general there will be many "natures" common to most of the sample, depending on the type and level of analysis one pursues: atomic, molecular, chemical, anatomic, genetic, social, etc. Which, if any, is relevant to fixing the range of correct application of the concept? Other things being equal, the relevant nature is the one that best accounts for subjective distance relations among instances of the concept.

Since animal species belong to taxonomic hierarchies, there will be many natures that all members of a species or subspecies share—natures associated with taxa of higher rank. All tigers share one nature in virtue of belonging to a kingdom, another in virtue of belonging to a phylum, and so on for class, order, family, genus, and species. Which shared nature is relevant to making a creature a tiger? It's plausible that the answer is determined by the role that the underlying natures play in accounting for the similarities among typical tigers—the perceptually salient features that distinguish tigers from non-tigers and give the observed sample its family resemblance.

There are in biology three competing taxonomic systems: the phenetic, the cladistic, and the evolutionary. The evolutionary is a taxonomy associated with a gradualist account of speciation, while the cladistic focuses on shared characteristics that derive from large evolutionary "leaps." Phenetic taxonomies are ahistorical, focusing on features in the phenotype. Which taxonomy is relevant to fixing the extension of the ordinary word 'tiger'? It's plausible that the answer is partly determined by the success of the various taxonomic schemes in giving an account of the perceptually salient features that distinguish tigers from non-tigers. For there would seem to be little *point* in imposing on thinkers norms of correct application in accordance with one taxonomic scheme that accounts less well than another for psychological distance relations in the sample.

From a microphysical point of view, variation in nucleon number (mass number) looms rather large; yet we count isotopes of an elemental substance as belonging to the substance. Isotopes of a substance occur naturally and have a different number of neutrons in the atomic nucleus than does the usual form of the element. Why do we not consider variation in neutron number to be relevant to the "shared nature" of the elemental sample, when we do so consider variation in proton number? Presumably because proton number plays a much larger role than neutron number in the account of perceptually salient features of the substance.

In many cases the shared nature will be social. Whether the "nature" is natural or social does not in general depend on whether thinkers

represent the kind as being natural or social. In the case of the concept *vegetable* it could be that everyone in the community thinks of the kind as biological, when in fact it is a grocery or culinary kind. Conversely, the extension of *armadillo* is fixed by the shared biological nature of the sample even if everyone thinks armadillos are artifacts. ("Who's making these armadillos?" they all wonder.) Explicit belief that the kind is natural or social counts for little in the face of the requirement that the relevant nature account for the psychological centrality and distance phenomena.

This last example illustrates a point due to Burge: general statements that are part of the "meaning" of a term may be challenged on the basis of perceptual ties to instances.[30] Even if meaning-giving criteria exist for the concept, they are not true trivially or by convention, but are open to challenge by competent speakers. It is possible to doubt whether sofas are pieces of furniture made or meant for sitting, even if one knows that this criterion is part of the meaning of the term 'sofa' in the language of one's community. The underlying intentional type is fixed in the first instance by its perceptually identified instances. The dialectic whereby meaning-giving characterizations may be doubted and tested without simply changing the subject is possible in virtue of shared exemplars, i.e., particular things perceived or remembered, or imagined things with the relevant salient properties, that are agreed to be "good examples" of the category.

What ought to be the principal motivating question for conceptual semantics? Not, I suggest, that of how meaning can arise in a physical world, but that of how normative semantic truths can hold given our psychological limitations. The question acquires force from a recognition of how inadequate are individualistically described response patterns for reconstructing intentional notions. We are good at such tasks as extracting central representations from a range of instances, but bad at such tasks as saying in fine detail what the range of correct application of our concept is, and our classificatory dispositions are fraught with error. Memory has a predilection for vividness and perceptual similarity, and is severely bounded in time, space, and experience. It ought to seem mysterious how creatures so limited can have concepts like that of *bird*, which correctly applies to all and only the birds, many of which we might not recognize as such. There is a gap between our psychological capacities, individualistically described, and our semantic reach.[31] We rely, of psychological necessity, on structured cooperation in the application of our concepts. The point of such structured cooperation is precisely to transcend the limitations of psychological distance, and so concept boundaries cannot be explained purely in terms of distance. But facts about psychological distance give point to semantic norms.

Arizona State University
Tempe, Arizona

NOTES

[1] See, for example, S. Kripke, *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980); H. Putnam, 'The Meaning of "Meaning",' in his *Mind, Language and Reality: Philosophical Papers Volume 2* (New York: Cambridge University Press, 1975), pp. 215–271; T. Burge, 'Individualism and the Mental', in P. French, T. Uehling, and H. Wettstein, eds., *Midwest Studies in Philosophy Volume 4: Studies in Metaphysics* (Minneapolis: University of Minnesota Press, 1979), pp. 73–121; and Burge, 'Other Bodies', in A. Woodfield, ed., *Thought and Object* (Oxford: Oxford University Press, 1982), pp. 97–120.

[2] For overviews of this literature, see E. Rosch, 'Principles of Categorization', in E. Rosch and B. Lloyd, eds., *Cognition and Categorization* (Hillsdale: Lawrence Erlbaum Associates, 1978), pp. 27–48, and E.E. Smith and D.L. Medin, *Categories and Concepts* (Cambridge, MA: Harvard University Press, 1981). An interesting collection of more recent work is U. Neisser, ed., *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization* (New York: Cambridge University Press, 1987).

[3] G. Rey's 'Concepts and Stereotypes', *Cognition* 15 (1983), pp. 237–262, is a fine exception. G. Lakoff, *Women, Fire, and Dangerous Things* (Chicago: University of Chicago Press, 1987) is an extended survey of linguistic and psychological data and a critique of what he calls "objectivist" metaphysics in philosophy.

[4] By 'everyday object category' I shall mean any category of perceptually recognizable objects or events for which most speakers have a monolexically expressed concept (e.g., *animal*, *woman*, *bird*, *bat*, *insect*, *fruit*, *carrot*, *elm*, *gold*, *salt*, *lake*, *star*, *wind*, *fight*, *kiss*, *dance*, *wedding*, *baby*, *school*, *tool*, *car*, *book*, *death*). Though I include artifactual and social categories, I exclude artificial categories constructed using Boolean operators (e.g., *red or square*), mathematical categories (e.g., *hexagon*, *prime number*), and terms introduced explicitly in the course of scientific theorizing (e.g., *caloric*, *quark*), all of which raise special issues that I shall not address.

[5] E. Rosch, 'Cognitive Representations of Semantic Categories', *Journal of Experimental Psychology: General* 104 (1975), pp. 192–233.

[6] E. Rosch, 'On the Internal Structure of Perceptual and Semantic Categories', in T.E. Moore, ed., *Cognitive Development and the Acquisition of Language* (Orlando: Academic Press, 1973), pp. 111–144.

[7] Rosch, 'Cognitive Representations of Semantic Categories', op. cit.

[8] L.J. Rips, E.J. Shoben, and E.E. Smith, 'Semantic Distance and the Verification of Semantic Relations', *Journal of Verbal Learning and Verbal Behavior* 12 (1973), pp. 1–20.

[9] E. Rosch and C.B. Mervis, 'Family Resemblances: Studies in the Internal Structure of Categories', *Cognitive Psychology* 7 (1975), pp. 573–605. Throughout this paper the points I make by reference to Rosch and Mervis's family resemblance measure apply equally to the more sophisticated measure in A. Tversky, 'Features of Similarity', *Psychological Review* 84 (1977), pp. 327–352.

[10] E. Rosch, C.B. Mervis, W.D. Gray, D.M. Johnson, and P. Boyes-Braem, 'Basic Objects in Natural Categories', *Cognitive Psychology* 8 (1976), pp. 382–439; J.A. Fodor, *The Modularity of Mind* (Cambridge, MA: MIT Press, 1983), pp. 94–97.

[11] E. Rosch, C. Simpson, and R.S. Miller, 'Structural Bases of Typicality Effects', *Journal of Experimental Psychology: Human Perception and Performance* 2 (1976), p. 491. The quotation does not represent Rosch's current view; see G. Lakoff, op. cit., pp. 39–57.

[12] L. Zadeh, 'Fuzzy Sets', *Information and Control* 8 (1965), pp. 338–353. For

semantically loaded glosses, see G. Lakoff, 'Hedges', *Journal of Philosophical Logic* 2 (1973), pp. 458–508; Rosch and Mervis, 'Family Resemblances', op. cit.; Rosch, 'Principles of Categorization', op. cit.

Lakoff, *Women, Fire, and Dangerous Things*, op. cit. pp. 43–44, claims that Rosch's 'Principles of Categorization' marks a shift in her views away from the assumption that typicality effects characterize the structure of the category as it is represented in the human mind. This is incorrect. In the article, Rosch is agnostic about the role of prototypes in models of processing, distinguishing theories about "the structure of categories from theories concerning the use of that structure in processing" (p. 36). She sees her work as constituting a theory of the structure of categories in "the perceived world and not a metaphysical world without a knower" (p. 29), and holds that "prototypes develop through the same principles such as maximization of . . . [intra-] category resemblance as those principles governing the formation of the categories themselves." (p. 37) It is precisely the view that the formation of categories can be explained in terms of intracategory resemblance with respect to perceived properties that I shall be criticizing.

[13] G. Rey, 'Concepts and Stereotypes', op. cit.

[14] D.N. Osherson and E.E. Smith, 'On the Adequacy of Prototype Theory as a Theory of Concepts', *Cognition* 9 (1981), pp. 35–58.

[15] Typicality effects can also be achieved with mathematical categories, such as *even number*, that are well defined. For example, a typical even number such as 4 is more quickly categorized than an atypical even number such as 13,572. See S.L. Armstrong, L. Gleitman, and H. Gleitman, 'What Some Concepts Might Not Be', *Cognition* 13 (1983), pp. 263–308.

[16] Osherson and Smith, op. cit., p. 37. It is clear from the surrounding context that the quoted remark correctly expresses their intent, but it misreads the formula. The conditional should be a biconditional; equivalently, $d(x, p) < d(y,p) \leftrightarrow c(y) < c(x)$. The characterization also oversimplifies in assuming a single prototype; this is particularly implausible (as án account of prototype theory) in the case of superordinate categories such as *tool* or *fruit*.

[17] R. Grandy, 'In Defense of Semantic Fields', in E. LePore, ed., *New Directions in Semantics* (Orlando: Academic Press, 1987), pp. 259–280.

[18] S. Kripke, *Naming and Necessity*, op. cit.; H. Putnam, 'The Meaning of "Meaning",' op. cit., especially p. 225; T. Burge, 'Other Bodies', op. cit.; T. Burge, 'Intellectual Norms and the Foundations of Mind', *Journal of Philosophy* 83 (1986), pp. 697–720; and T. Burge, 'Wherein is Language Social?', forthcoming.

[19] It is therefore not surprising that formal measures of family resemblance that have been proposed in the literature generally presuppose a binary category. $F(a,C)$ is a measure of a relation between an instance and the binary category C. Each psychologically salient feature is, in effect, weighted according to the number of instances of C that share it, and in this computation each instance of C counts for one.

[20] Here I am indebted to Tyler Burge; see especially his 'Individualism and Psychology', *Philosophical Review* 95 (1986), pp. 3–45.

[21] Cf. T. Burge, "Wherein is Language Social?", forthcoming.

[22] See, for example, Osherson and Smith, op. cit., p. 57.

[23] My view has affinities to that of G. Rey, 'Concepts and Stereotypes: A Reply to Smith, Medin, and Rips', *Cognition* 19 (1985), pp. 297–303. Rey argues against the core vs. identification procedure view, and for the psychological significance of his 'Hypothesis of External Definitions': "[essential] defining conditions [of the concept's extension] are provided by the optimal account of the concept, and so may not be known to competent users of it, whose representations of those conditions would in those cases be marked by empty slots."

I am skeptical, however, about the theoretical utility of both the notions "optimal account" and "empty slot". No one in the subject's community may be in possession of the relevant account, as Rey acknowledges. If the relevant account is supposed to be

part of an ideally completed science, then the reliance on such an ideal account is rather attenuated—not at all like reliance on others in the community for access to the sample itself. For example, there is no causal connection to the ideal optimal account or possessors of it. Moreover, if the relevant account is actually known, the subject may through stubbornness or honest error fail to recognize it as such. In any case the subject may have his own strongly held but mistaken theory about the underlying nature. The postulated "empty slot" need not exist. (My account also differs from Rey's in that I tie my anti-individualistic account of concepts to an account of the nature of psychological explanation).

[24] Further aspects of this background commitment are explored in chapter 6 of my *Individualism and the Cognitive Sciences*, unpublished PhD dissertation, UCLA, 1986.

My arguments for a background commitment to an underlying nature appeal to (I think) obvious facts about higher cognition, but some recent psychological literature can be seen as supporting my view. Psychologists have tried to articulate broad structural principles governing the construction of object categories. For example, it has been suggested that categories are constructed so as to maximize "cue validity". The central idea is that humans construct categories so as to organize the correlated salient features of environmental objects, maximizing within-category similarity while minimizing between-category similarity. But it is notoriously difficult to make this idea precise. For "cue validity", see E. Rosch and C.B. Mervis, 'Family Resemblances', op. cit.; for critical discussion, see G.L. Murphy and D.L. Medin, 'The Role of Theories in Conceptual Coherence', *Psychological Review* 92 (1985), pp. 289–316.

I conjecture that no approach to this issue can succeed that treats salient features as merely statistically correlated while ignoring the role of "underlying natures" as causes of the feature correlations. Some recent psychological work, especially that of D.L. Medin, takes categories to be organized around *theories* rather than feature correlations *per se*. For discussion, see the Murphy and Medin article cited above, and also D.L. Medin and W.D. Wattenmaker, 'Category Cohesiveness, Theories, and Cognitive Archeology', in Neisser, op. cit.

For example, in certain categorization tasks, experimental subjects preferred sorting by properties that could be seen as not merely correlated but causally connected; see D.L. Medin, W.D. Wattenmaker, and S.E. Hampson, 'Family Resemblance, Conceptual Cohesiveness, and Category Construction', *Cognitive Psychology* 19 (1987), pp. 242–279. I leave for another occasion the task of arguing in detail that the new emphasis on theories, and the experimental phenomena that have been taken to support it, are best interpreted in light of my views in this paper.

[25] See the papers in D. Kahneman, P. Slovic, and A. Tversky, eds., *Judgment Under Uncertainty: Heuristics and Biases* (New York: Cambridge University Press, 1982).

[26] See the discussion of "cognitive archeology" in D.L. Medin and W.D. Wattenmaker, 'Category Cohesiveness, Theories, and Cognitive Archeology', in Neisser, op. cit., pp. 25–62.

[27] See R. Grandy, 'Information-Based Epistemology, Ecological Epistemology and Epistemology Naturalized', *Synthese* 70 (1987), pp. 191–203.

[28] W. Bechtel, 'Connectionism and Rules and Representation Systems: Are They Compatible?', *Philosophical Psychology* 1 (1988), pp. 5–16. See also note 3 of G. Rey, 'Sanity Surrounded by Madness', *Behavioral and Brain Sciences* 11 (1988), pp. 48–50.

[29] R. Grandy, 'Information-Based Epistemology, Ecological Epistemology and Epistemology Naturalized', op. cit., p. 201.

[30] T. Burge, 'Intellectual Norms and the Foundations of Mind', op. cit.

[31] Compare what Chomsky calls "Plato's Problem"; quoting Russell, he writes: "How comes it that human beings, whose contacts with the world are brief and personal and limited, are nevertheless able to know as much as they do know?" N. Chomsky, *Knowledge of Language: Its Nature, Origin, and Use* (New York: Praeger, 1986) p. xxv.