A reassessment of the shift from the classical theory of concepts to prototype theory

Eric Margolis

Department of Philosophy, Rutgers University, Davison Hall, Douglass Campus, PO Box 270, New Brunswick, NJ 08903-0270, USA

Abstract

A standard view within psychology is that there have been two important shifts in the study of concepts and that each has led to some improvements. The first shift was from the classical theory of concepts to probabilistic theories, the most popular of which is prototype theory. The second shift was from probabilistic theories to theory-based theories. In this article, I take exception with the view that the first shift has led to any kind of advance. I argue that the main reasons given for preferring prototype theory over the classical theory are flawed and that prototype theory suffers some of the same problems that have been thought to challenge the classical theory.

1. Introduction

The psychological study of human concepts has had a rich history in the last twenty years, with a series of widely endorsed theories and a pervading sentiment that progress has been made. In a recent review of the literature, Douglas Medin writes that there have been two important shifts in theories of concepts (Medin, 1989). The first shift, due largely to the work of Eleanor Rosch, was from the classical theory of concepts to probabilistic theories. The second shift was from

probabilistic theories to, as Medin puts it, theory-based theories. Medin clearly thinks that in both cases the psychology of concepts has improved with the change and now defends theory-based theories (Murphy & Medin, 1985). Medin's sense of things, if not universal, is certainly mainstream. At the same time, it is unwarranted. In my view, the historical sequence Medin cites traces no progression; the reasons typically cited to prefer one theory over the other are all flawed, and the problems that infect earlier theories crop up in their successors. In this article, I focus on the shift from the classical theory to probabilistic theories, sticking to the most popular version of the latter – prototype theory. I argue that the main reasons commonly given for preferring prototype theory to the classical theory are no good.

Theories of concepts are often put in ways that obscure their psychological content, so they are open to a certain amount of interpretation. After some preparatory remarks in section 2, I begin with a standard yet inadequate characterization of the classical theory and prototype theory. In sections 4 and 5, I consider two revisions – the most plausible interpretations of the literature. Since, on either interpretation, the arguments for preferring prototype theory fail, I conclude that the received view in psychology is wrong.

2. The representational theory of mind

There is probably no univocal notion of a concept in psychology or the cognitive sciences. Different researchers are likely to mean different things when they defend a theory of concepts. Nonetheless, in the part of psychology with which we will be concerned, certain unifying assumptions are commonly made, most importantly those that go with the representational theory of mind (RTM). RTM is an account of the nature of mental states and processes. In brief, RTM claims that having a propositional attitude involves bearing one of a set of particular functional relations to a representation and that mental processes are typically causal interactions amongst representations. Within cognitive psychology, this picture has been refined so that mental processes are viewed as *computational* processes defined over mental representations (e.g., Fodor, 1975). I take it that the general picture is familiar, but its consequences for the study of concepts are not always made explicit. In particular, it is natural to assume, given RTM, that concepts are constituents of thoughts. The idea is that thoughts have something akin to syntactic structure, allowing the representational system to admit of a compositional semantics (e.g., Fodor & Pylyshyn, 1988), where concepts are to be identified with a subset of the representations from which thoughts are composed. Thoughts, on this view, are complex mental representations with satisfaction conditions, that is, mental sentences. Concepts are the subsentential representations that constitute thoughts. Note that this is a different

use of the term "concept" than one often finds in philosophical discussions. Philosophers tend to think of concepts as abstract objects, the semantical values of open sentences. For a philosopher, the concept cat is what the expression "cat" means, perhaps the property *cat* or the set of cats.[1] Psychologists are not at all hostile to this view, only psychologists tend to think that the relation natural language expressions bear to concepts, in the philosopher's sense of the term, is mediated by mental representations, which psychologists call "concepts". For present purposes, I will stick to psychological usage.

If concepts are understood as representations, then a theory of concepts amounts to a theory of representations, at least the part of such a theory that concerns subsentential representations. And while it is doubtful that there is any widespread agreement about what a complete theory of representations should look like, two areas of interest are clear, both of which will be important later. The first concerns the semantical properties of representations; the second concerns the nature of conceptual structure.

Representations have formal and semantic properties. To some, it is their semantic properties that are most puzzling. The concept CAT refers to cats. On the present view, having the concept CAT involves having a mental representation, a symbol, presumably encoded in the brain, which refers to cats. But how is it that neurologically realized symbols refer? In virtue of what does your cat-representation pick out the set of cats? Philosophers have studied these questions, in one form or another, since the inception of RTM in the seventeenth century. The prevailing options have been that representations have their semantical properties either by virtue of resemblance relations (Berkeley, Hume) or by virtue of causal relations (Locke). Neither approach has fared well, but in recent years causal accounts have improved with advances in information-based semantics (e.g., Dretske, 1981). Moreover, theories based on natural selection have offered some new possibilities (e.g., Millikan, 1984). In any case, one clear project for a theory of concepts is to address the question of how it is that representations refer.

A related question concerns the compositional structure of concepts per se. In general, the semantical value of a complex concept will be a function of the semantical values of its constituents taken with their manner of organization. This is just the principle that the representational system admits of a compositional semantics. But there may be, in addition, theoretically significant constraints on the compositional structure of concepts. Empiricists in philosophy and almost everyone in cognitive psychology and artificial intelligence tend to think that this is the case for lexical concepts, those expressible by monomorphemic terms, such

as the concepts CAR and BIRD. A standard empiricist view, for example, is that the primitive constituents of lexical concepts express sensory qualities and that, as a result, all concepts reduce to sensory ones. Suffice it to say for now that with any commitment to compositionality comes a research program which, in part, seeks constraints on the assignment of constituents to complex concepts.

3. Confusing concepts with categories

Let us now turn to the classical theory and its successor, prototype theory. In this section, I wish to emphasize a difficulty in assessing the theories. The problem is that they are often put in terms of *non-psychological* relations, in particular, in terms of non-psychological relations that hold amongst the semantic values of concepts – categories. Concepts and categories are easy to confuse, a kind of use/mention error. Concepts are psychological constructs and, we are to suppose, mental representations with semantic properties. Categories, on the other hand, are what concepts are about. They are the groups of objects and events and so on that representations represent.[2] The concept CAT, for example, refers to cats; or, if you like, CAT has the set of cats as its semantic value; or, perhaps, CAT expresses the property *cat*, which picks out the set of cats. In any event, the difference between the concept CAT and cats should be obvious and as glaring as the difference between the word "cat" and cats. The point is that, in the first instance, psychologists should focus on the concept, not the category. But there is a tendency in psychology to get this wrong or at least to obscure the difference to the extent that the relevant psychological theories seem to concern categories exclusively.

Consider the following formulation of prototype theory by an early advocate, James Hampton (Hampton, 1981, p. 149):

> A polymorphous concept [Hampton's term for one that conforms to prototype theory] is one in which an instance belongs to a certain category if and only if it possesses a sufficient number of a set of features, none of which need be common to all category members. For example, "sweetness" is a feature of fruit, but several important members of the category (such as lemons) do not possess it. Thus there may be many features that are neither necessary (required for

[2]"Objects" is used loosely here to include abstract objects, such as properties and sets. Sometimes psychologists are concerned with a special subset of these and reserve the term "category" for just this set. Elizabeth Shipley distinguishes arbitrary classes of objects from categories, where "categories", according to her terminology, are classes of objects that have the following properties: "(a) they have labels that are used to identify objects; (b) they serve as the range of inductive inferences; and (c) their members are believed to share a "deep resemblance" (Shipley, in press). For my purposes, this distinction can be ignored, and I will continue to use the term "category" in the broader sense.

membership) nor sufficient (guaranteeing membership) but yet are important to the category definition (other examples are "growing on trees", "roundness", and "having peel").

On this characterization prototype theory concerns the extra-psychological conditions for category membership. The difference between the classical theory and prototype theory amounts to the difference between (1) and (2):

(1) *The classical theory*
   All instances of a category share a set of properties singly necessary and jointly sufficient for membership within the category.

(2) *Prototype theory*
   Category membership is a matter of having some sufficiently many properties that members of the category tend to have.

Notice that the theories as expressed in (1) and (2) have *no* psychological implications. Both are straightforward metaphysical theses, the kind that should be argued on a case-by-case basis across the special sciences. Do all birds share some set of properties singly necessary and jointly sufficient for being a bird or is the *category* birdhood more lenient? In principle this could turn out to be a question for psychologists, but without strong philosophical arguments to the contrary, it would seem to be a matter for the zoologists to settle. Moreover, I doubt many psychologists would be willing to endorse the sorts of philosophical considerations that would allow psychological theories to arbitrate questions about non-psychological phenomena.[3]

4. First revision of the classical theory and prototype theory: conditions for having a concept

Here is where we are. Psychologists tend to think that prototype theory is an improvement upon the classical theory. What is more, while the two are clearly meant to be understood as psychological theories, they are often put in a way that leaves them without psychological import. A common formulation of the theories, given in (1) and (2), cannot be right. Before we can decide whether prototype theory does indeed improve upon the classical theory, we need to reformulate the theories. We need to frame them in terms that make explicit their intended psychological content. The rest of this paper explores two possibilities. I am not sure whether either of these is exactly what the leading psychologists have had in mind, but they strike me as the most plausible interpretations of the literature. Since I think that on either interpretation prototype theory is no better than the classical theory, I conclude that the received view in psychology is wrong.

---

[3] For an exception, see Lakoff (1987), especially the preface and the second part of Book I.

4.1. The revision

My first proposal for interpreting the two theories in a psychologically relevant way is to view them as building upon (1) and (2) by adding further conditions for *having a concept*, conditions that exploit the non-psychological theses embedded in (1) and (2). The conditions that come most readily to mind are epistemological:

(3) *The classical theory*
    (a) All instances of a category share a set of properties singly necessary and jointly sufficient for membership within the category.
    (b) Having a concept involves knowing the conditions of membership within the corresponding category.

(4) *Prototype theory*
    (a) Category membership is a matter of having some sufficiently many properties that members of the category tend to have.
    (b) Having a concept involves knowing the conditions of membership within the corresponding category.

On this characterization of the theories, their principal aim is to explain the semantical properties of representations. (Recall that concepts are being identified with representations.) In brief, the theories are versions of the description theory of reference, according to which a representation refers to what it does by virtue of being associated with a particular description that picks out its referent (for an extended discussion of description theories, see Devitt & Sterelny, 1987, Ch. 3). The relevant descriptions are those that specify the correct extra-psychological conditions for category membership. (3) and (4) differ only in what they take these conditions to be, that is, with respect to (3a) and (4a). Moreover, (3a) and (4a) are identical to (1) and (2), respectively.

To get a feel for the theories, consider an example that appears to support the first. Grandmothers, we all know, are female parents of at least one parent. With some plausibility, *grandmother* satisfies (3a): there is a set of properties that are necessary and sufficient for membership within the class of grandmothers, namely, being female and being a parent of a parent. This much is about the category only, not the concept, and satisfies our original version of the classical theory. (3), however, is only partly satisfied. It says something about the concept GRANDMOTHER as well. According to (3), having the concept GRAND-MOTHER involves knowing that grandmothers are female parents of a parent. It does seem natural to think that having GRANDMOTHER involves knowing at least this much about grandmothers. Would you not hesitate to say that someone

had the concept GRANDMOTHER, say a young child, if she did not know that grandmothers were parents?

4.2. A questionable assumption

The trouble with the concept GRANDMOTHER is that in important respects it is atypical. In many cases, people do not know the conditions for category membership. This is part of the point of Kripke's and Putnam's work on natural kind terms (Kripke, 1972; Putnam, 1970, 1975). Kripke and Putnam both think that competence with a natural kind term cannot require knowledge of the conditions of the term's application (conditions for category membership) because in so many cases where we clearly are semantically competent we do not know the conditions (sometimes out of ignorance, sometimes out of error). If they are right, then what they say about natural kind terms goes, mutatis mutandis, for natural kind concepts. And if what they say extends to various other types of terms – which clearly Putnam thinks is the case – then that too goes, mutatis mutandis, for the relevant kinds of concepts. I take it that their examples need no reviewing, but I would like to review the structure of their argumentation. It bears a resemblance to a line of argument that has been used against the classical theory. The difference is that, in psychology, the argument has been taken to apply uniquely to the classical theory, where, in the context of Kripke's and Putnam's discussions, it is clear that the argument would extend to prototype theory as well. Let me explain.

Many psychologists have been intrigued by the fact that people can rarely specify necessary and sufficient conditions for category membership, especially since, despite their failures, they tend to insist that concepts have definitions (McNamara & Sternberg, 1983). This has suggested to some that the classical theory must be wrong, yet the tendency has been to maintain much of its spirit. In particular, the second part of the classical theory, (3b), is often kept intact. The idea is that having a concept involves knowing the conditions of membership within the corresponding category – that the classical theory had this much right – only the conditions that must be known are less strict. They are less strict because the conditions required for category membership are less strict. Objects need not instantiate each and every property of a stringent set of properties for category membership. Rather, they need only instantiate some number of them. To have a concept, then, requires knowing less – in a sense. The reason I qualify this remark is that in one important respect (3) and (4) make equally strong claims about the epistemic status of ordinary people, advertisements notwithstanding. Both require that people be privy to the real extra-psychological conditions for category membership – (3b) and (4b). That is, both require that people be highly knowledgeable in the way that Kripke's and Putnam's examples

suggest that we are not. If Kripke and Putnam are right, then the relevant fact is not that people cannot in general specify necessary and sufficient conditions for category membership. It is that people cannot in general specify the conditions for category membership *whatever they are*. This suggests that the problem with the classical theory, as stated in (3), is its claim about what is needed for having a concept, (3b), not its claim about category membership, (3a). Then prototype theory, as stated in (4), is no better off; it makes the same claim about what is needed for having a concept – (3b) and (4b) are identical.

Suppose you had a prior commitment to (3) and you discovered that agents cannot in general specify necessary and sufficient conditions for category membership, that is, that (3b) is false if (3a) is true. Then you would have two immediate options: one, save (3b) and drop (3a); two, save (3a) and drop (3b). The route I have been attributing to some prototype theorists is the former. But it is important to see the latter is an option. Kripke and Putnam do not explicitly discuss concepts, but, as far as natural kind terms go, they do maintain something akin to (3a) and try to tell a new story about semantic properties of terms in which causal connections to the world play the significant role and knowledge is of secondary importance. Putnam does require that people have a socially given "stereotype" for most natural kind terms. A stereotype provides a specification of some of the more noticeable surface-level aspects of a kind (e.g., that lemons are yellow); however, the properties given by the stereotype are not taken to specify conditions for category membership, even partly. One might have thought that since Kripke and Putnam and prototype theorists all agree that people cannot provide necessary and sufficient conditions for membership within many categories that they all agree that the classical theory is wrong and, moreover, that Kripke and Putnam have provided grounds for thinking that the classical theory is wrong, helping the prototype theorists. Such an assessment would be misleading. To the degree that Kripke and Putnam have reasons for thinking that the classical theory is wrong they have reasons for thinking that prototype theory is wrong. What is more, their resulting theories bear an affinity to the classical theory, if either, since both accept a form of (3a). As things stand, we have been given no reason for preferring prototype theory.

The objection I am pushing is related to ones made earlier by Georges Rey in an exchange with Edward Smith, Douglas Medin and Lance Rips (Rey, 1983; Rey, 1985; Smith, Medin, & Rips, 1984). The exchange starts with Rey's review of Smith and Medin's influential book, *Categories and concepts*. Rey notes that Smith and Medin tend to assume throughout their discussion that the conditions for having a concept are epistemological, that two people have the *same* concept just in case they have the same beliefs about members of the corresponding category, most notably beliefs about how to determine whether an arbitrary object is a member of the category. Rey thinks that this sort of condition for conceptual individuation is hopeless for a variety of reasons. First, people are

generally prepared to revise their beliefs about the conditions for category membership and about how to tell whether something belongs to a particular category. If we accept the epistemological construal of conceptual individuation, then "any two people who use different procedures, or one person who uses different procedures at different times, would ipso facto have different concepts" (Rey 1983, p. 249). Second, as Kripke and Putnam show, people are often ignorant or wrong about the conditions for category membership and how to determine whether an object is a member of a category. But you cannot explain conceptual identity in terms of beliefs that people do not have.

While I am sympathetic with these arguments, neither is quite the one I have in mind, though the difference may be a matter of emphasis. Recall that, for the moment, we are construing the classical theory and prototype theory as aiming to explain the semantical properties of representations. The proposal we have been considering is that they do this by granting people beliefs which suffice to determine (in the metaphysical sense) the correct extra-psychological conditions for category membership. Rey objects that people often do not have the beliefs to serve this purpose and that insofar as they do, these beliefs are not sufficiently stable to guarantee that different people (or the same person over time) will have the same concepts. The point I wish to emphasize – which I think is largely unrecognized in the literature – is that if it is an argument against the classical theory that people do not have the relevant beliefs, then it is *also* an argument against prototype theory: (3) and (4) agree that having a concept involves knowing the conditions for membership within the corresponding category. In fact, not only does the problem affect both the classical theory and prototype theory (as they are expressed in (2) and (3)) but it extends to what Medin calls theory-based theories of concepts. These too share the epistemological clause. The chief difference between theory-based theories and the others is that the beliefs specified by the former are supposed to express causal/explanatory relations. This difference is insignificant, since theory-based theories buy into the crucial assumption that people are knowledgeable in ways that they often are not.

5. Second revision of the classical theory and prototype theory: the compositional structure of lexical concepts

So far, I have looked at two ways of understanding both the classical theory of concepts and prototype theory. Neither squares with the common sentiment in psychology that prototype theory is better than the classical theory. If the theories are to be understood according to (1) and (2), then they have no psychological import. If, on the other hand, they are to be understood according to (3) and (4), there is no reason for thinking that prototype theory fares better than the classical theory. In this section, I would like to look at another natural interpretation of

the theories, this time viewing them as theories that constrain the assignment of constituents to lexical concepts. I will argue that on this interpretation prototype theory is still no better than the classical theory.

5.1. The revision

Many psychologists – as well as many linguists and philosophers – think that lexical concepts have constituent structure in exactly the same sense that patently complex concepts do. In other words, they think that lexical concepts tend to be constructed out of simpler concepts, some of which may themselves be lexical (e.g., Clark, 1973; Miller & Johnson-Laird, 1976). If we take this proposal seriously, a number of research programs emerge. One concerns the discovery of the conceptual primitives for humans. Another concerns the discovery of the structure for particular lexical concepts or classes of lexical concepts. A plausible way of understanding the classical theory and prototype theory is that they are intended to address this last issue. The theories on the present formulation continue to have two parts, the first of which addresses the conditions for category membership. Previously, the second part of each theory specified how it is that representations refer to what they do. Now the second part of each theory issues constraints on the constituent structure of lexical concepts:

(5) *The classical theory*
   (a) All instances of a category share a set of properties singly necessary and jointly sufficient for membership within the category.
   (b) Conceptual structure reflects category structure in that a lexical concept, C, is composed of component concepts, $c_1, c_2, \ldots, c_n$, such that c_1–c_n are concepts of those properties that determine membership in the category that C is a concept of

(6) *Prototype theory*
   (a) Category membership is a matter of having some sufficiently many properties that members of the category tend to have.
   (b) Conceptual structure reflects category structure in that a lexical concept, C, is composed of component concepts, $c_1, c_2, \ldots, c_n$, such that c_1–c_n are concepts of those properties that determine membership in the category that C is a concept of.

Concepts, we are supposing, are representations, and complex concepts are complex representations composed of simpler ones such that the semantical value of a complex is a function of the semantical values of its parts taken with their manner of organization. (5) and (6) differ on a subset of the complex concepts. Consider the concept BLACK CAT. BLACK CAT is presumably complex,

composed of the simpler (though, not necessarily simple) concepts BLACK and CAT. The semantical value of the concept BLACK CAT (the category *black cat*) is a function of the semantical values of its immediate constituents (*black* for BLACK, *cat* for CAT). On the present construal of the theories, both agree on this much. And both agree that there is nothing further to say about the structure of non-lexical concepts.[4] They differ only on the lexicals. We are to think of questions concerning the structure of lexical concepts as part of a research program. (5) and (6) are competing approaches within this research program. According to (5), the constituents of a lexical concept will express properties that are singly necessary and jointly sufficient for membership within the category expressed by the lexical concept. According to (6) the constituents of a lexical concept will express properties for which having some number is sufficient to be a member of the category expressed by the lexical concept. For the rest of this section, I take a critical look at two of the leading arguments for favoring prototype theory over the classical theory, when they are understood along these lines. I argue that neither supports (6) over (5). Since there are no other plausible interpretations of the theories, it is fair to conclude that the received opinion in psychology is mistaken: prototype theory is not an advance upon the classical theory.

5.2. Conceptual fuzziness

One argument that is generally understood to support prototype theory over the classical theory is, in brief, that the classical theory cannot account for the fuzzy nature of concepts and that prototype theory can. In much of psychology, it is simply taken as a datum that concepts are fuzzy. It is not always clear whether the fact in question is supposed to concern categories or concepts, whether, say, it is the set of bird that is supposed to be fuzzy or the concept BIRD. The two may go together. If concepts themselves are supposed to be fuzzy, this might be a function of the fuzziness of the corresponding categories. Let us assume – tentatively anyway – that categories are in fact fuzzy. The present alleged objection to the classical theory is that (i) concepts are typically fuzzy and (ii) the classical theory cannot handle fuzziness. Douglas Medin cites an argument of this sort as one of the leading contributors to the demise of the classical theory. "The classical view", he says, "implies a procedure for unambiguously determining category membership; that is, check for defining features. Yet there are numerous cases in which it is not clear whether an example belongs to a category" (Medin, 1989, p. 1470). He adds, "It is . . . easy to see that the probabilistic view

---

[4]This is not to say, however, that both theories are equally able to account for the compositional structure of non-lexical concepts (Fodor, 1981, p. 292-297; Osherson & Smith, 1981).

[prototype theory] may lead to unclear cases. Any one example may have several typical properties of a category but not so many that it clearly qualifies for category membership" (p. 1471). Thus, it appears, we have an argument for (ii): if category membership is a matter of satisfying each and every property of a set of properties then no sense is to be made of degrees of satisfaction, that is, either an object is a member of a category or not. In contrast, if category membership is a matter of satisfying a high number of a set of properties then objects will be members to the degree that they satisfy this set (objects which have all of the properties will be members to the highest degree, objects which have none will be non-members, and objects which have a number around the threshold for membership will be members to a low degree). Explanations of this sort have moved many psychologists towards prototype theory. However, the explanation is no good.

Prototype theory is supposed to explain the fuzziness of concepts by its commitment to a metaphysical thesis whereby categories admit of degrees of membership. One problem with this explanation is that while it might suffice to explain conceptual fuzziness, we have been given no reason to think that conceptual fuzziness cannot be explained consistent with the classical theory. The classical theory, we are told, entails that membership within a category is all or nothing, hence that concepts are not fuzzy. But why exactly is membership all or nothing? The answer is supposed to be that this falls right out of the classical theory's conditions for category membership. According to the classical theory, the requirement for membership is that an object have each and every property of a particular set. So all members of a category are equal in the sense that as members of the category they have the same properties. From this, we are supposed to conclude that membership fails to come in degrees. But the inference does not go through. At the very least, there is a relevant and clear sense in which the classical theory does admit of degrees of membership, namely, through the potential fuzziness of the properties required for membership. Consider a simple example, *black cat*. In one sense, the extension of the category *black cat* is perfectly determinate: something is a black cat just in case it is black and a cat (has the properties *black* and *cat*). At the same time, it does not follow that for every object it is either a black cat or not (full stop). *Black cat* may admit of degrees so long as either *black* does or *cat* does, and I doubt there is much dispute about the former. The same goes for any complex concept – including lexicals – however its semantics is to be projected from its constituents. Hence, fuzziness does not in itself argue against the classical theory.

In a way this understates the classical theory's wherewithal because the response concedes that category fuzziness is just that, a question of the degree to which a target is a member of a category. But it is not at all clear that the fuzziness so often taken to be a datum reflects anything about category member-ship. In certain cases I am sure that we do think membership is graded. In other

cases, however, the fuzziness we note is purely epistemological. That is, given what we know, we find it more or less difficult to make judgements of category membership, all the while supposing that membership is an absolute matter. Imagine coming across penguins or whales for the first time. We have all been educated to know that penguins are birds and that whales are mammals. Yet it is easy enough to put ourselves in an earlier frame of mind, where people were uneasy about calling penguins birds and thought whales were fish. Still, the working assumption then as now seems to be that if Xs are birds/mammals then Xs are absolutely birds/mammals (i.e., 100% bird/mammal). Our uneasiness never turned on the supposition that penguins might be birds to a small degree, a smaller degree than, say, robins are birds. It is just that we thought that penguins were less obviously birds, if indeed they were birds (which they are), because they do not fly. As Georges Rey comments (Rey, 1983, p. 248):

> [A] distinction needs to be drawn between two sorts of "unclear cases": those, like that of *euglena* . . . , which may well be on the borderline between animal and plant: and those, like that of *tomatoes* . . . , which may be (metaphysically) clear cases of fruit (check the dictionary!), even though people may be (epistemologically) confused about them.

Our intuitions that categories are fuzzy sometimes reflects nothing more than our own limited epistemic situations. And to the degree than these intuitions depend on a belief that membership within a category is genuinely graded, they do not favor prototype theory over the classical theory.

5.3. Categorization

We have just looked at one argument that is ordinarily taken to support prototype theory over the classical theory and we found it lacking. I would like to end by looking at one other argument, one which many psychologists think demonstrates that prototype theory is the proper successor to the classical theory (e.g., Hampton, 1981). The argument appeals to facts about categorization.

Categorization, in brief, is the psychological process whereby people make judgements about whether an object is in the extension of a given category. If you were to respond to the query "Is a robin a bird?", the process underlying your ability to answer the question would be a categorization process. Categorization is itself an interesting object of study. One might wonder how it is that we are able to make reliable judgements about category membership. The general framework for answering this question has been, for the most part, representationalist. The idea is that, in categorization, subjects compare their representation of the probe (in this case ROBIN) to their representation of the queried category (BIRD). Theories of categorization compete according to how they fill in the details. For example, the classical theory of concepts, as it is put in (5), suggests that one

compares the probe representation (ROBIN) to see if it has all of the constituents of the queried-category representation (BIRD). If ROBIN lacks any of BIRD's constituents, then the answer is "No"; if ROBIN has all of BIRD's constituents, then the answer is "Yes". The reason (5) suggests this sort of model is because according to (5) category membership requires having every property of a proprietary set. It seems reasonable, then, that in *deciding* whether an object is a member of a category one should determine whether it has each of the relevant properties.

If anyone once took this particular model of categorization seriously, no one does today. The reason is that it does not accommodate a robust set of categorization data – typicality effects (for the classic review of the literature, see Smith & Medin, 1981; for a recent and critical review, see Barsalou, 1987). Typicality effects are data concerning the speed with which categorization judgements are made. The generalization seems to be that subjects categorize more "typical" members of a category faster than they categorize less "typical" ones, where the typicality measure of the members of a category is independently established.[5] One way to yield a typicality measure is to have subjects rank members of a category on a scale of, say, one to seven, for how good they are as examples of the category. Subjects find this to be a natural task. For example, American college students tend to rank robins higher than penguins as examples of the category *bird*. Correspondingly, they answer "Is a robin a bird?" faster than they do "Is a penguin a bird?" This correlation, between typicality measure and categorization time, is a paradigmatic typicality effect. Now because typicality effects are robust, one would think that a theory of categorization should accommodate them. For this reason, many psychologists have had doubts about the classical theory. Nothing in the model of categorization that the classical theory suggests predicts that more typical members of a category will be categorized faster. And, at the same time, prototype theory, as put in (6), can account for the data, so long as categorization is viewed as a similarity comparison process and the following relatively innocuous assumption is made: that probes with a high typicality measure are more similar to the queried category than probes with a low typicality measure are. Typicality effects are then explained as a measure of the similarity of the target representation to the category representation. The idea, in short, is that subjects can compute the comparison between ROBIN and BIRD faster than they can compute the comparison between PENGUIN and BIRD because ROBIN shares more of relevant constituents of BIRD than PENGUIN does. What is more, the adoption of a similarity comparison model of categorization is not at all arbitrary given prototype theory, since, as far as category membership goes, the theory requires that members of a

[5] Generally, the data range over correct positive categorizations, for example, that you say "Yes" to "Is and X a Y?" when Xs are Ys.

category have some number of properties; it does not require that they have all of the properties, or any particular ones, of a specific set. In sum, typicality effects are easy to account for given prototype theory and difficult to account for given the classical theory. Hence, we are to believe that prototype theory is the more promising.

The problem with this line of reasoning is that the data are equivocal. There is strong evidence that typicality effects occur in categorization judgements even in the special cases where it can be demonstrated that subjects know that the relevant category can be defined in terms of necessary and sufficient conditions (Armstrong, Gleitman, & Gleitman, 1983). That is, in a few special cases where it is very unlikely that prototype theory could be true, typicality effects still occur. Thus it is unclear whether anything special about prototype theory is what generates the typicality effects in other cases. Moreover, there is nothing in the classical theory that requires categorization to operate on the constituent structure of representations. Categorization might, for example, exploit clear non-constitutive relations amongst concepts. This is the basic idea behind two-component theories of categorization. On these theories, a concept comes with an identification procedure that has little to do with known conditions for category membership. One may know that grandmothers are female parents of a parent and yet, for purposes of categorization, judge whether someone is a grandmother according to whether she exhibits stereotypical properties of a grandmother, for example, grey hair and glasses (Landau, 1982; Rey, 1983). In any event, insofar as typicality effects are supposed to argue against the classical theory, the data are equivocal. This result has not been fully appreciated in the cognitive sciences. Many psychologists continue to insist that typicality effects reflect an advantage of prototype theory over the classical theory. Ellen Markman, for example, who studies the acquisition of natural language terms, assumes that typicality effects count as good evidence that lexical concepts have what she calls a family resemblance structure – they conform to prototype theory (Markman, 1989, Ch. 3). She says this despite reviewing Armstrong et al.'s data, with nothing of substance to reconcile the obvious tension. But, to repeat, Armstrong et al. have established that inferences from categorization data to theories of conceptual structure are suspect. So far as we know, prototype theory has *no* advantage over the classical theory.

6. Conclusion

I started, in section 3, with a common formulation of the classical theory and prototype theory. The problem with this formulation is that it strips the theories of psychological import. With this in mind, I proposed two revisions. According to the first, the theories aim to explain how it is that representations refer to what

they do. Under this interpretation, however, prototype theory suffers much the same difficulty as the classical theory, that people often do not know the conditions for category membership. According to my second revision, the theories are theories of the constituent structure of lexical concepts. But under this interpretation too there is no reason to prefer prototype theory to the classical theory. Arguments that cite conceptual fuzziness and typicality effects are unconvincing. They rely on equivocal data. The principal arguments that many have taken to support prototype theory over the classical theory are no good. I conclude that, for all we know, prototype theory offers no advantage over the classical theory; the shift from the classical theory to prototype theory has been unwarranted. What is more, understanding why this shift has been unwarranted is especially important if the psychological study of concepts is to advance. Some of the problems that infect the classical theory persist in the literature, largely because the classical theory has been abandoned for the wrong reasons.

References

Armstrong, S., Gleitman, L., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*, 263–308.

Barsalou, L. (1987). The instability of graded structure: implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: ecological and intellectual factors in categorization* (pp. 101–140). New York: Cambridge University Press.

Clark, E. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. Moore, (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.

Devitt, M., & Sterelny, K. (1987). *Language and reality: an introduction to the philosophy of language*. Cambridge, MA: MIT Press.

Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

Fodor, J. (1975). *The language of thought*. New York: Thomas Y. Crowell.

Fodor, J. (1981). The current status of the innateness controversy. In *Representations: philosophical essays on the foundations of cognitive science* (pp. 257–316). Cambridge, MA: MIT Press.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: critical analysis. *Cognition*, *28*, 3–71.

Hampton, J. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition*, *9*, no. 2, 149–156.

Kripke, S. (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.

Landau, B. (1982). Will the real grandmother please stand up? The psychological reality of dual meaning representations. *Journal of Psycholinguistic Research*, *11*, no. 9, 47–62.

McNamara, T., & Sternberg, R. (1983). Mental models of word meaning. *Journal of Verbal Learning and Verbal Behavior*, *22*, 449–474.

Markman, E. (1989). *Categorization and naming in children: problems of induction*. Cambridge, MA: MIT Press.

Medin, D. (1989). Concepts and conceptual structure. *American Psychologist*, *44*, no. 12, 1469–1481.

Miller, G. & Johnson-Laird, P. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Millikan, R. (1984). *Language, thought, and other biological categories: new foundations for realism.* Cambridge, MA: MIT Press.

Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* no. 3, 289–316.

Osherson, D., & Smith, E. (1981). On the adequacy of prototype theory as a theory of concepts." *Cognition, 9,* 35–58.

Putnam, H. (1970). Is semantics possible? In H. Keifer, & M. Munitz (Eds.), *Language, belief and metaphysics.* New York: State University of New York Press.

Putnam, H. (1975). The meaning of "meaning". In K. Gunderson, (Ed.), *Minnesota studies in the philosophy of science, VII.* Minneapolis: University of Minnesota Press.

Rey, G. (1983). Concepts and stereotypes. *Cognition, 15,* 237–262.

Rey, G. (1985). Concepts and conceptions: a reply to Smith, Medin, and Rips. *Cognition, 19,* 297–303.

Shipley, E.F. (in press). Categories, hierarchies, and induction. In D. Medin, (Ed.), *The psychology of learning and motivation: Volume* 30. *Representation and processing of categories and concepts.* Orlando, FL: Academic Press.

Smith, E. & Medin, D. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Smith, E., Medin, D., & Rips, L. (1984). A psychological approach to concepts: comments on Rey's "concepts and stereotypes". *Cognition, 17,* 265–274.