# The Role of Similarity in Natural Categorization

James A. Hampton
Department of Psychology
City University

Chapter in: M.Ramscar, U.Hahn, E.Cambouropolos, & H.Pain (Eds.) 2000. *Similarity and categorization.* Cambridge: Cambridge University Press.

The intuitive idea that those things that we put things into categories because we find them similar appears to be non-controversial, if not circular. Cars are clearly more similar to other cars than they are to trees, and trees more similar to other trees than they are to cars. However, a number of theorists have recently questioned the degree to which the notion of similarity is sufficiently clearly defined and constrained to serve as an explanation of our categorization. In this chapter, I discuss the arguments for and against basing categorization on a notion of similarity, and conclude that, construed broadly, similarity may yet be the best explanation of how most of our conceptual categories function. I propose a distinction between concepts viewed as a cultural phenomenon and concepts at the psychological level, and suggest a naive model of conceptual development that starts with concepts as similarity clusters and only in restricted cases replaces these with more causal or theory-based conceptual representations.

**Similarity-Based Categorization**

What is the evidence that similarity plays a role in categorization? To answer this question, as Sloman and Malt (this volume) point out, we need to be quite precise about what we mean by similarity. We form categories of many different kinds in the course of everyday cognition, and it could be claimed that they are *all* based on similarity. But this would be to render the notion so broad as to be empty. For example, Barsalou (1983) pointed to the existence of what he termed *ad hoc* categories such as Birthday Presents for Your Mother, or Things to Take on a Camping Trip. Members of these categories are of course similar in one important respect -- things to take on a camping trip are all similar in as much as they are all good things to have along when camping. But this tautological similarity does not go far in explaining how this category is constructed. Nor does it appear that the degree to which something is a good member of the category is related in any way to its similarity to other members in any respect other than its property of being in the category.

Another class of categories which could only tautologically be explained in terms of similarity is the class of concepts with *explicit* definitions. For example being a Triangle depends on a small number of explicit criteria, such that only similarity *in those respects* is relevant to class membership. To say that all triangles are similar to each other in respect of being closed figures having three straight lines for sides, three angles, and internal angles that sum to 180° is to say little more than that all triangles possess all these properties. Similarity reduces to identity. Categories of this kind are clearly *not* based on similarity, except in a tautological sense. Similarity must mean more than simple identity on a particular dimension or combination of dimensions.

By contrast, we form many other categories, many of them stable and long-term parts of our conceptual repertoire, which *do* show a strong link to similarity. These categories are characterized by having *no explicit definition* (unlike ad hoc categories or explicitly defined categories), a number of associated properties which are *generally* true of category members, although not universally so, and a graded structure such that some items are more clearly and uncontroversially members of the category than are others. Rosch and Mervis (1975) termed these concepts Prototype Concepts. Prototypes are ideal or central tendencies around which categories form. The category is then composed of all items that are sufficiently similar to the prototype (for a formal treatment see Hampton 1995a). According to Prototype theory, our biological inheritance and social and cultural environment provide the dimensions along which we note similarity and difference. Where a number of these dimensions correlate in our experience, then a category of similar items is formed, to which we give a name, and which we can then use as a concept in our thinking and language. Once the dimensions have been determined, clustering of the world into classes is relatively automatic. Indeed there are advanced statistical theories of how items may be clustered based on partially correlated dimensions (van Mechelen et al., 1993).

It may be countered (e.g. Fodor, 1998) that

this theory of similarity-based concepts requires there first to be non-similarity-based features -- that is that similarity at the macro-level depends on identity at the micro-level. If such is the case, then it could be argued that our prototype concepts are really just rule-based concepts where the criteria for membership are (a) disjunctive and logically complex, and (b) implicit and so unavailable to conscious report. There are two rejoinders to this view.

First, it is not a necessary requirement on micro-features or dimensions that they involve simple binary truth values (i.e. identity versus non-identity). Most of the putative dimensions of which prototypes are considered to be constructed (such as visual shape and size, functionality, origin) may have graded degrees of match themselves. Second, rule-based categorization does not generate in any straightforward manner predictions of either gradations in typicality or vagueness in categorization, unless one assumes that not only are the rules for categorization hidden from conscious report, but that they are also unreliable in their operation. Complexity of the rules *per se* is insufficient to account for the fact that the boundaries of our categories are so ill-defined.

The process of evolving similarity-based prototypes can also involve recursion. In order to obtain a cleaner and more generally useful set of categories, we may adjust the weights of dimensions, and even construct new dimensions from which to build our concepts. We don't construct conceptual categories merely to satisfy some drive for neatness -- they play a central role in everyday behaviour and action, they permit predictive inference, they are a necessary building block for acquiring and using knowledge of the world. Concepts evolve in order to maximise their general utility value, according to some (as yet unknown) criterion of utility (for one suggestion see Pothos and Chater, this volume).

It is at this point in the story that a number of psychologists have argued that something other than mere similarity defined over features must be playing a role. Part of our drive for knowledge and understanding is the search to replace similarity-based clusters by explicitly defined concepts with broad explanatory power. Keil (1989) refers to this as the principle of "original sim" -- that children's initial concepts are based on pure similarity, which is then replaced in time with deeper, more theory-like kinds of conceptual understanding.

The progress of science is a testimony to just this process. One of the first tasks of a natural scientist is to construct a typology of the domain. What are the relevant kinds of thing in the domain about which the scientific narrative can be told? A botanist or zoologist may therefore first classify on the basis of gross physical similarity -- trees with similar leaf shape, branching structure, trunk markings are more likely than not to belong to the same natural kinds. From this beginning, the scientist then may wish to discover interesting facts about the tentative kinds. It is at this stage that initial categories may be refined and individuals reclassified to make a clearer story. Whales are no longer considered fish on the basis of habitat and external features, when a mass of detail of their internal physiology and their behaviour (air breathing, giving birth to live young and suckling them) becomes known. In this case, classification based on "surface similarity" (by which we generally mean similarity in respect of the most immediately available information) is replaced with classification based on deeper similarity. It remains an interesting question to what extent this process reflects a different type of categorization rather than reflecting the same categorization process applied to a wider range of information. One could argue that the biologist classifies a whale as a mammal rather than a fish, because when all that is known about mammals, fish, and whales is taken into account, the similarity to mammals is much greater -- even before differential weighting is given to "deep" as opposed to "surface" features (Ahn and Dennis, this volume).

To take a second example, when medical research first identifies a phenomenon it defines a *syndrome* -- a cluster of symptoms, and conditions of occurrence, with some predictive value in terms of treatment and prognosis. (Most mental illnesses are still at this stage of understanding.) It is characteristic of syndromes that cases may be more or less typical, and more or less clear members of the syndrome. Frequently cases may arise that are borderline to the syndrome, possessing some similarity to the prototype, but not enough to be clearly identifiable as an example. Discovery of an aetiology linked to the syndrome -- such as an infectious organism, or physiological/biochemical malfunction -- will usually allow the syndrome to be replaced by a clearly defined disease/condition category, with its own set of diagnostic tests. Ahn (1999) has evidence of the bias to select causal features as central features of categories (see also the chapter in this volume by Ahn and Dennis).

Note that the set of patients and their

symptoms has not changed -- the world has not become more clear-cut in any way. However whereas before a case was borderline because it showed marginal levels of similarity to other cases, a case will now be borderline if the critical diagnostic tests do not come out with a clear answer. There is a shift from an uncertainty which is *conceptual* in its origin, to an uncertainty which is *epistemological* -- that is to say that a case is now borderline because we cannot discover clearly enough whether the defining agent is at work. Our uncertainty has to do with our state of knowledge, rather than our state of understanding.

This analogy with science serves as a template for the debate that followed publication of Murphy and Medin's (1985) attack on similarity as a basis for natural concepts. Biologists adopt particular taxonomies because they permit explanatory accounts of evolution or because they capture similarity defined across the broadest range of features. Physicians seek to *explain* the presenting symptoms through a causal account. In an analogous fashion, Murphy and Medin argued that we use our concepts as ways of explaining the world to ourselves and others. It then follows that we categorize not on the basis of a similarity cluster, but on the basis of selecting the concept that best explains the instance to be categorized. This alternative account of categorization has also had wide acceptance in the developmental field (Carey, 1985, Keil, 1989).

The difference between similarity and explanation-based or "causal theory" accounts of categorization was brought into sharp focus in a paper by Rips (1989). Rips attacked the unconstrained nature of similarity as a basis for categorization, and reported a number of demonstrations of cases where the similarity account clearly fails. Each of these demonstrations involved the discovery of a dissociation between similarity and categorization. If categories are formed around prototypes, then it should not be the case that one item could be more similar to, or more typical of the category than another, but yet less likely to belong. In formal terms, this means that there should be a monotonic function relating similarity to a category and membership in that category. Rips provided three cases where this constraint was broken.

**The pizza-coin example**

In his first case, subjects were asked to consider a hypothetical item that was exactly half way between two categories, one a fixed category and the other a variable category. For example they had to imagine a circular object that was half way in diameter between the largest US quarter they had seen and the smallest pizza they had seen. Subjects then judged whether this object was either (a) more similar to or typical of one category rather than the other, or (b) more likely to be a member of one category rather than the other. Rips reported a dissociation between similarity and typicality on the one hand, where people generally considered similarity to be about equal to each category, and likelihood of membership on the other hand where people generally judged the object more likely to be in the variable category (the pizza in this case).

This demonstration is only counter-evidence to the prototype theory of concepts if one assumes that the function relating similarity in diameter to categorization probability is equivalent for the two concepts. (There have also been problems in replicating the main result, Smith & Sloman 1994.) I have argued (Hampton, 1979, 1995b) that a prototype concept must involve three aspects: a conceptual representation of the class, a similarity metric for determining how similar an instance or subclass is to the prototype, and a threshold placed on similarity to determine whether to include the instance or subclass in the category. If we simplify the pizza and coin example just to the dimension of size (as Rips asked his subjects to do), and if we grant that similarity is simply a matter of absolute difference in diameter, even then it is clearly not the case that either the feature weight attached to diameter, or the thresholds for categorization will be the same for the two conceptual categories. It is in the nature of the fixed categories that Rips selected that the weight attached to size, and the similarity threshold for categorization is very high. No doubt in the real world the diameter of quarters or tennis balls is variable (they can only be manufactured to a certain tolerance, and are then subject to wear), but it is also clear that the variability is much less than that for pizzas or watermelons. The contribution of a mismatch on diameter to the likelihood of categorization is very different in the two cases. One might argue (with Rips) that this is just the point - that the theory-based account can explain <u>why</u> there are these differences in the weight and threshold for similarity in diameter between the fixed and variable cases, whereas the prototype account cannot. However it is not necessary to invoke deeper causal reasoning to explain the differences. Simple observation of quarters and pizzas will lead to a representation of the distribution of sizes in

each case, and so to the fixing of feature weights and threshold values. It is true that on this account prototypes must be capable of representing not just the central tendency of the class but also the variability along each dimension. However there is nothing in the prototype model to rule this out (see Hampton, 1995b, 1998), and if prototypes are poor at representing distributions of values, the other main class of similarity-based categorization models -- exemplar models (e.g. Nosofsky, 1988) -- are ideally suited to representing both the mode and the range of variability of exemplars of a class.

The clearest demonstration that knowledge of variability does not imply causal theories is an experiment by Fried and Holyoak (1984) in which two abstract visual categories were learned by subjects (as the work of two abstract painters). When one category was arranged to be more variable than the other, then items that were equal in distance from the two categories were more reliably categorized as belonging to the more variable category.

**Bimodal and skewed distributions**

The second example of a dissociation between similarity and categorization was reported in a paper by Rips and Collins (1993). Subjects were given information about the shapes of two (non-normal) distributions of values on some dimension - for example daily maximum temperatures for two particular locations. They were then given particular values and asked to judge their similarity or typicality as an example of each distribution, or asked to say which distribution the item was more likely to belong to. Under these conditions, people tended to base similarity judgments on distance from some measure of central tendency. Likelihood of categorization however was based on the probability density corresponding to that point of the distribution. For example a temperature of 55 degrees Farhenheit might reflect the annual average temperature, but because of strong seasonal variation, the most common temperatures might be 35 degrees (winter) and 75 degrees (summer). Rips and Collins found that people would say a temperature of 55 degrees was more similar to the distribution, but was less likely to have come from it, whereas a temperature of 75 was less similar but more likely to have come from the distribution.

It is arguable how easy it is for people to interpret a request to judge how similar one number is to those in a distribution. From the results, it appears that the subjects adopted some measure of distance from the central tendency as defined by the mean or median. For categorization, they quite correctly looked at the relative likelihood of a particular value in each of the two distributions, and judged probability of belonging on that basis. Interestingly, if Rips and Collins' data is examined more critically it appears that judgments of typicality were performed on the same basis as categorization, and were not driven by distance from central tendency.

What sense can be made of a differentiation between similarity and typicality? Are they not expected to always be equivalent? The answer is that they are not. First, similarity will tend to be calculated across whatever features or dimensions are considered relevant to a particular context of comparison (Medin, Goldstone and Gentner, 1993). It is therefore quite feasible for similarity to differ from typicality if the context is not clearly one of determining categorization. In the Rips and Collins experiments, it is clear that subjects rejected basing typicality on distance from the central tendency, and instead used frequency as a basis for judging typicality. There is supporting evidence from Barsalou (1985) that frequency of instantiation can have an effect on typicality, and when similarity is held constant, as in well-defined concepts, frequency may even be the main determinant of typicality (Armstrong, Gleitman & Gleitman, 1983). It is also important to note that categorization should ideally take account of both similarity and frequency information. A diagnostician must combine the typicality of the pattern of symptoms observed with knowledge of the rarity of a condition in arriving at a most likely diagnosis.

**Transformations and metamorphoses**

Rips' final example involved a creature (or artifact) which metamorphosed from one category into another. For example, in one scenario, a bird-like creature was transformed into an insect-like creature through an environmental accident. When asked whether the creature that changed was more similar to or typical of a bird or an insect, people went for the insect category. However they also judged the creature (marginally) more likely to be a bird. In a variant on this scenario, the subjects were told that the metamorphosis was the result of natural maturational processes (of the kind that turns caterpillars into butterflies or tadpoles into frogs). In this case, subjects judged that the immature form (before the change) was more similar to and typical of a bird, but that the creature

was more likely to be an insect.

These data are more convincing, showing as they do a failure to categorize along the lines of typicality, and that the manipulation of the causal story of how the change came about has an effect on categorization. Zachary Estes and I (Estes & Hampton, forthcoming) decided to follow up Rips' experiments in order to test the robustness of his findings. In particular we were concerned about a number of aspects of the procedure and the results. First, the procedure was not completely standardized as between the two causes of change, with variations in whether different names were given to the two phases of the creature's life, and whether a question was asked about the creature in general, or just about its initial phase. Second, we were concerned that subjects were presented with a booklet in which beneath each scenario all three questions were asked - which category is it more similar to, which category is it more typical of, and which category is it more likely to belong to. We felt it unlikely that subjects would feel happy about going through the whole booklet giving identical responses to each of these scales. If you ask three questions, there is a strong pragmatic expectation that you are looking for different answers. Finally we noted that whereas the similarity and typicality ratings were fairly clearly biassed towards one or other category, the categorization data were suspiciously close to the 50% level, suggesting either that subjects were evenly divided in their opinions, or that they were responding randomly across scenarios, having no clear basis on which to make their decision.

We conducted three experiments, using animal scenarios similar to those employed by Rips, and with a number of new controls built in to the procedure. Instead of a control condition in which no transformation occurred, we asked subjects to judge the creature either at the start of the story, or at the end of the story. It was therefore possible for subjects to express the anti-essentialist belief that a creature began as one kind, and then turned into another kind. In the first experiment we chose to treat the question to be asked as a between-subjects factor, so different groups of subjects judged typicality and categorization. In the final experiment we reinstated Rips' procedure of having subjects rate both scales at the same time for each scenario. The results were quite clear. When only asked to judge one scale, there was no dissocation between judgments of typicality and judgments of categorization. Unlike Rips, we found that the large majority of our subjects responded as "phenomenalists", in effect deciding to place a creature in whichever category it was more typical of. When the bird was transformed into an insect, in their view it became an insect, even though its offspring were bird-like again. On the other hand, when subjects in the final experiment were required to make both judgements together about each scenario, the dissociation was restored, with typicality following appearance, and categorization hovering around 50%. Even in this experiment however, a close examination of individual subjects' patterns of responding showed that many individuals were still responding as phenomenalists.

Perhaps the clearest message to take from the instability of Rips' dissociation results is that people find it difficult to make decisions about categorization in these counterfactual worlds in which creatures are capable of changing from one kind into another. It is for this reason that they show such a strong effect of the pragmatic expectations built into the procedure. In any event, the case for dissociation of typicality and categorization must be considered unproven.

**Evidence for Similarity in Categorization**

In the light of these various critiques of similarity-based categorization it is worth briefly reviewing the evidence *for* similarity-based categorization of the kind proposed by prototype or exemplar models.

First there is the *fuzziness* of many of our concepts. When asked to reflect on the meaning of words like "fish", "art", or "sport", people find it very hard to give a theoretically satisfactory account of the underlying concepts. They are however very good at generating ways in which members of the category differ from other things in the same domain. They can also quickly recall or create examples to illustrate what a typical category member might be. There is apparently a rich source of semantic information associated with the concept, but it does not appear to be organized in anything like the neat structures proposed by the causal theory view. The lack of organization and internal coherence becomes particularly clear when people's reasoning with concepts has been studied. Hampton (1982) showed that people may quite willingly agree (for example) that School Furniture is a type of Furniture, and that a blackboard is a type of School Furniture, but yet disallow that a blackboard is a type of Furniture. Categorization was not treated as a universally transitive relation, in contradiction of both classical and even fuzzy

logic (Zadeh, 1965). Instead, I argued that each separate category judgment was made on the basis of similarity. As the basis on which similarity changes between the two judgments, it is then quite possible to obtain intransitive categorizations.

Tversky and Kahneman (1983) found similar effects on subjective probability judgments. They found that people used similarity to prototype as a means of judging subjective likelihood, even when this strategy produced clearly illogical results, such as judging it more likely that a radical female student would have become a feminist bank teller, than that she would simply have become a bank teller. This conjunction fallacy was paralleled by the finding of overextension of conjunctive categories by Hampton (1988, 1996). People were willing to say for example that Chess was a Sport which is a Game, even though they had earlier judged that Chess was not a Sport. Hampton (1997) replicated this result with a between-subjects design, and extended the demonstration of inconsistent classification to the case of negation. For example 80% of participants in one group considered Tree Houses to be Buildings, yet 100% of participants in another group considered them to be Dwellings that are *not* Buildings. Our conceptual categories display a degree of flexibility and context sensitivity which is much more easily captured by a similarity-based process than by a fixed theoretical schema. A recent study by Sloman (1998) is a further demonstration of how similarity can be shown to affect people's conceptual reasoning. In one demonstration, Sloman found that people were more likely to accept the truth of a logically necessary conclusion when the terms of the two premises were similar than when they were not. Similarity effects are pervasive in people's attempts to reason logically, and a very simple explanation for this finding is that our conceptual system is heavily dependent on similarity-based conceptual processes.

A critical test of similarity-based categorization is the extent to which categorization can be influenced by "irrelevant" kinds of similarity. There is a distinction in the literature, originally introduced by Smith, Shoben and Rips (1974), between Defining and Characteristic Features. It was their notion that there were many properties of objects which might determine how typical they were of their class, but which would be irrelevant to their category membership. Their example was that the ability to fly is very typical of birds, and so flying birds are more typical members of their class. Flight as such however is irrelevant to determining whether a creature is a bird or not, since there are both birds that do not fly and other creatures (notably insects) that do fly. Smith et al. termed this idea the Characteristic Feature Hypothesis. Hampton (1995b) set out to test whether Characteristic Features (CF) are in fact always irrelevant to categorization in practice. To test this idea, I created sets of six hypothetical objects for each of a number of concepts. Each object either possessed or lacked a full set of CF. In addition each object either had a full set of Defining Features (DF+), lacked at least one Defining Feature [DF-), or had a *partial match* to the Defining Features [DF?]. The aim of the experiment was first to show that when the object possessed the DF, categorization would be clearly positive, and when it lacked at least one DF, then it would be clearly negative, regardless of the CF. The critical test was then to be whether the CF would affect categorization when the DF were only partially matched. For example consider an object which *partially* matched the DF of umbrellas - it was designed to keep things from falling on you, but instead of protecting you from the rain it was intended to protect you from acorns and twigs when picnicking under a tree. Would this odd object be more likely to be categorized as an umbrella if it had the classical domed shape and material of umbrellas, than if it was built in some different shape and material?

In the event this critical second test could not easily be performed. The reason was that it proved very hard (even after four replications of the experiment with improved materials and improved instructions), to find CF which did not still influence categorization, even when the DF were clearly present or absent. For example one example of DF+, CF- was the following description:

"The offspring of two zebras, this creature was given a special experimental nutritional diet during development. It now looks and behaves just like a horse, with a uniform brown color."

When asked if this was really a zebra, only a third of the subjects agreed, the rest ignoring the genotype in favor of the phenotype, contrary to the assumptions of both biological theory and psychological essentialism (see the results of the Estes and Hampton study reported earlier). Similar problems occurred when I attempted to pit the intended function of artifacts (assumed to reflect their real nature) against their outward appearance. People tended to be influenced by similarity along dimensions which logical analysis suggests should be irrelevant -- *unless* of course categorization is

based on similarity calculated across a wide range of dimensions. (Malt and Johnson, 1992, reached similar conclusions about category membership of artifact concepts not being solely based on function).

**Does Categorization Depend only on Typicality?**

According to the Prototype Model, categorization proceeds by assessing the similarity of an instance or subclass to the concept prototype, and then testing whether it passes some threshold criterion for category membership. If this model is inadequate, then as Rips (1989) argued, it should be possible to demonstrate non-monotonicity between measures of similarity to prototype (on the one hand) and likelihood of category membership (on the other). Non-monotonicity implies that while instance A may be more typical of a category than instance B, when it comes to categorizing them, B is more likely to be categorized in the category than is A.

Hampton (1998) set out to discover to what extent non-monotonicity of this kind could be found in everyday common semantic categories. Rips (1989) used a variety of unusual examples to dissociate similarity and categorization, and it is questionable how generalizable such results are to the more usual process of deciding if subclass A is a member of category B. It is therefore interesting to know whether categorization in a common category such as Fish or Vehicle follows typicality in the category, or whether dissociations between the measures can be found. To answer this question, I reanalyzed a data set published in 1978 by McCloskey and Glucksberg, in which they had two groups of subjects making judgments about 18 semantic categories. One group were asked to make typicality judgments for a list of 30 items for each category, ranging from clear category members to clear non-members. A second group gave a simple Yes/No categorization decision about each item for each category. This second group returned a month later and made their categorization decisions a second time. McCloskey and Glucksberg (1978) found that the categorizations showed fuzziness in two respects. First, there was considerable disagreement amongst people over which items should be included in the categories and which should not. This disagreement was reflected in a large number of items with Categorization Probability at intermediate levels between 0 and 1. Second, there was a considerable degree of within-subject inconsistency when the follow-up test was made. High levels of disagreement and inconsistency were most noticeable for items in the *middle* of the typicality scale -- that is for items that were neither clear members nor clear non-members. McCloskey and Glucksberg concluded that categorization in many semantic categories is fuzzy, rather than all-or-none, and that there is a considerable amount of instability in how we categorize.

The data from this research were published as an Appendix, and provided an opportunity to test for non-monotonicity directly. Typicality ratings are *prima facie* direct measures of how similar an instance or class is to the category prototype (assuming there are no marked differences in frequency of instantiation). The instructions for typicality emphasize that a high rating should be given to items that are *representative* or *good examples* of the class as a whole. On the other hand Categorization Probability is a simple way of measuring the degree to which something is categorized in a class. If we assume that there are random and individual sources of variation in categorization, then the group measure of how many subjects say X is in category Y may be taken as a fairly direct measure of the degree to which X is considered to belong in Y by each individual.

The data were therefore analyzed in order to examine the mathematical relationship between mean rated typicality and categorization probability. The first conclusion was that there were clear differences between individual categories in terms of how clearly categorization probability could be predicted from typicality. While some categories, such as Sport, gave a very clear monotonically rising threshold function, with practically no systematic deviation, for others such as Fish there was a considerable spread of items above and below the threshold function, and plenty of evidence for non-monotonicity. There was no link however between how well the measures correlated and the kind of semantic domain. There were good and bad fits in both natural kind and artifact categories.

In order to explore the various possible reasons why some items should not follow a clean threshold function such as that shown for Sport in Figure 1, but instead should be scattered above and below the function as in the case of Fish, a regression function was fitted to the data from all 17 categories, (one category was excluded for technical reasons), and the residual categorization probability was calculated for each item. The items with categorization probability significantly higher

or lower than that expected for their typicality were examined in more detail, and a number of hypotheses suggested themselves to account for the variation. First, there were a number of very unfamiliar items such as Euglena, or Lamprey, which had categorization probability higher than expected from Typicality. Typicality ratings are known to be affected by familiarity (Barsalou, 1985; Hampton & Gardiner, 1983). It is therefore quite likely that low familiarity with an item may depress its Typicality without affecting its categorization.

On the other hand there were items with lower categorization probability than expected, which appeared to be semantically associated with the category, but not actually category members. Examples were Orange Juice as a Fruit, or Egg as an Animal. Bassok and Medin (1997) have shown that semantic associatedness can give a sense of similarity, and it is not unreasonable to suppose that Typicality ratings may also reflect associatedness to an extent that is not seen in categorization itself.

Two further hypotheses were related to the distinction that Rips, Keil and others have stressed -- namely the distinction between the surface appearance of objects, and their deeper nature. Some items bear a superficial resemblance to a category to which they do not belong -- a whale as a Fish is perhaps the best known example. Other items bear little resemblance to the category to which they *do* belong -- as might be the case for tomatoes and Fruit. It may be expected that items that are *technically not members* should have lower category probability than expected, while those with are *only technically members* should have higher probability than expected.

A final hypothesis concerned the effect of contrast categories on typicality and categorization. Similarity to a prototype may be calculated without regard to any contrasting or overlapping categories of which the item may be a member. Categorization however may proceed in a more contrastive manner, in that people may prefer to categorize each item in just one category (as in the *mutual exclusivity principle*, adopted by young children in word learning -- Clark, 1973). If an item is a better member of some contrasting or overlapping category, then perhaps its categorization probability would be less than expected from its typicality.

These various hypotheses were collected together and tested by collecting judgments from a new sample of US students for each item concerning its Unfamiliarity, the degree to which it was Only Technically a member, or Technically Not a member, the degree to which it was judged a Part or Property rather than a true member, and the degree to which it also belonged in a Contrast category. These five new variables were entered into a regression to predict residual categorization probability when the effect of Typicality had been removed. Four of the five variables proved to be significant predictors, in the expected direction. Items that were Unfamiliar, or were Only Technically members, were associated with positive residuals -- they were more likely to be categorized positively than warranted by their typicality. Items that were associated parts or properties, or that were Technically Not members were associated with negative residuals -- they were less likely to be categorized positively than was warranted by their typicality. The Contrast variable had no overall predictive effect on residual categorization probability.

A subsequent analysis compared the 4 biological categories (Animal, Bird, Fish and Insect), with the 5 artifact categories (Clothing, Furniture, Kitchen Utensil, Ship and Vehicle). It was found that the two "Technical" predictors were significant for the biological categories, but not for the artifacts. On the other hand, the Contrast category predictor was significant only for the artifact categories. This difference is consistent with the fact that people may be influenced by biological classification in the zoological categories, but that no corresponding theory exists for artifacts. Similarly, artifacts often fall into overlapping categories (a knife may be either a tool, a weapon or a kitchen utensil), whereas biological categories are usually mutually exclusive. Further evidence of differences in the function relating categorization to typicality for biological kinds and artifacts has been reported by Diesendruck and Gelman (1999).

Hampton (1998) concluded that there were few systematic deviations from monotonicity and many of them could be accounted for by the effects of familiarity or associatedness on typicality ratings. There was also evidence that typicality gives less weight to "technical" or deeper aspects of objects than does categorization, most particularly in biological categories. This conclusion fits with Ahn and Denis' view that deeper "causal" features are more heavily weighted in categorization tasks (Ahn and Denis, this volume).

**What Role Does Similarity Play?**

In this chapter I have reviewed arguments and evidence that similarity-based categorization is

in fact a widespread phenomenon, affecting not only the common everyday use of categories, but also people's reasoning processes about those categories. It would probably be foolish to argue that all our categories are constructed on the basis of putting similar things together. We would certainly have made little progress culturally or scientifically if our conceptual repertoire were limited to such categories. How then can the evidence for similarity-based categorization be squared with this widely held notion that our concepts should *not* be based on similarity?

I propose that the difficulties of squaring the evidence for similarity-based categorization with the strong theoretical intuition that concepts must be based on more than "mere similarity" can be resolved by noting a distinction that is rarely made in the literature - that between concepts as cultural constructions and concepts as elements of mental representation. By a cultural construction, I mean the concept that a culture has developed and evolved over many generations of thought and discovery, and which represents the "received" or correct understanding of the world at any particular moment in the evolution of a culture. In most cultures there will be particular experts with socially validated responsibility for learning these concepts from the previous generation, for reflecting and/or improving upon them and for passing them on to the next generation. In so-called primitive societies they may be the elders who tell the myths of the ancestors, or who keep the mysteries of some religious cult. Since the Enlightenment in Western civilisation, an increasingly large group of these experts have been involved in the development of scientific understanding of the world. Society is so structured that most users of a conceptual term such as "lemon" or "bird" have little or no knowledge about the biological theory underlying the concepts of species, and so they are happy to defer to the expert. This deference will however be much greater in the case where some social value attaches to the categorization. The subjects in my experiments were apparently willing to ignore biological essence (e.g. parenting or offspring) in determining whether a creature was really a zebra or not. After all, little hangs on this question for the average student. However it is clear that questions such as whether a piece of paper is a £20 note, whether a gemstone is a diamond, or whether a painting is a real Van Gogh, cannot be decided with sufficient reliability by the lay person and require deference to experts. Similarity to a prototype is insufficient in such cases. The critical information that is needed for categorization involves tracking the banknote back to its origin in the mint, or testing the stone for its hardness and refractive index, or proving a provenance that shows the history of the painting since it left the studio of the painter.

In the more common case where little depends on the "correctness" of a categorization with respect to the cultural norm, then each individual may be using a somewhat different schema for representing the concept, and may defend his or her right to consider it to be correct. The psychological question therefore becomes one of determining what are the mental representations of concepts that people use in every day life. Given most people's deep ignorance of most domains of knowledge, one has to conclude that an over-emphasis on theory-based concepts may seriously overestimate the conceptual sophistication of the mentally represented concepts that psychologists explore in their experiments.

Of course, even if we commonly use similarity as the basis for categorization, we also have the capacity to think in a more precise logical fashion. We can define explicit terms such as Prime Number or Triangle, or we can define explicit goals to be satisfied (as in Barsalou's ad hoc categories). This type of axiomatic thought is fundamental to the success of mathematics and the mathematical sciences, and by its nature it makes little use of similarity. Scientific concepts tend to form all-or-none categories, which can enter into logical relations and scientific laws to derive deductive proofs. However psychological studies of this type of thought have found is that it is actually very *difficult* for most people. School teachers have to spend hours and hours of patient explanation to get the majority of students to understand the principles of mathematics or scientific laws and their concepts, and the majority of the population never succeed in mastering the necessary skills in more than a rudimentary form. From the earliest days of experimental psychology it has been shown that people are poor at following the abstract logic of syllogisms, conditionals, or probability. They are also poor at using analogy in problem solving unless surface similarity helps to cue the appropriate connection. Arguments that similarity-based categorization is inadequate since it cannot form a solid foundation of concepts for logic and reasoning are therefore founded on a dubious premise -- namely that most people have such a foundation readily available to them. It is perhaps more realistic to suppose that similarity

forms the basis of most people's concepts most of the time, and that some individuals, with a lot of training and with the advantage of the cultural transmission of ideas from great thinkers of the past are able to develop more advanced thinking skills in particular domains. Dimly remembered lessons may lead us to believe that our concepts are clearer than they really are -- or to defer to experts as keepers of the truth. However for everyday purposes we are content to continue putting together things that are (superficially or deeply) similar. After all, such a system serves us perfectly well for most daily purposes.

**Acknowledgements**

The author acknowledges the support of the British Academy, the Nuffield Foundation, the University of Chicago and the French Ministry for Higher Education and Science. This research has benefited from discussion with many colleagues over the years, notably Larry Barsalou, Daniele Dubois, Zachary Estes, John Gardiner, Dedre Gentner, Douglas Medin, Gregory Murphy, Lance Rips, Brian Ross, and Steven Sloman.

**References**

Ahn, W. (1999). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition, 69,* 135-178.

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

Barsalou, L.W. (1983). Ad hoc categories. *Memory and Cognition*, *11*, 211-227.

Barsalou, L.W. (1985). Ideals, Central Tendency, and Frequency of Instantiation as Determinants of Graded Structure in Categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 629-654.

Bassok, M., & Medin, D.L. (1997). Birds of a feather flock together: Similarity judgments with semantically rich stimuli. *Journal of Memory and Language*, *36*, 311-336.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Clark, E.V. (1973). Meanings and Concepts. In J.H.Flavell, & E.M.Markman (Eds.), *Handbook of child psychology*: *Vol. 3. Cognitive development* (pp 787-840). New York: Wiley.

Estes, Z., & Hampton, J.A. (1999). Similarity and Essentialism in Categorization of Natural Kinds. Unpublished manuscript.

Fried, L.S., & Holyoak, K.J. (1984). Induction of Category Distributions - A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 234-257.

Fodor, J.A. (1998). *Concepts: where cognitive science went wrong*. Oxford: OUP.

Hampton, J.A. (1979). Polymorphous Concepts in Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 441-461.

Hampton, J.A. (1982). A Demonstration of Intransitivity in Natural Categories. *Cognition*, *12*, 151-164.

Hampton, J.A. (1988). Overextension of conjunctive concepts: Evidence for a Unitary Model of Concept Typicality and Class Inclusion. *Journal of Experimental Psychology*: *Learning, Memory and Cognition*, *14*, 12-32.

Hampton, J.A. (1995a). Similarity-based categorization: the development of prototype theory. *Psychological Belgica*, *35*, 103-125.

Hampton, J.A. (1995b). Testing Prototype Theory of Concepts. *Journal of Memory and Language*, *34*, 686-708.

Hampton, J.A. (1996). Conjunctions of Visually-based Categories: Overextension and Compensation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 378-396.

Hampton, J.A. (1997). Conceptual Combination: Conjunction and Negation of Natural Concepts. *Memory and Cognition, 25*, 888-909.

Hampton, J.A. (1998) Similarity-based Categorization and the Fuzziness of Natural Categories. *Cognition, 65,* 137-165

Hampton, J.A., & Gardiner, M.M. (1983). Measures of Internal Category Structure: a correlational analysis of normative data. *British Journal of Psychology*, *74*, 491-516.

Keil, F.C. (1989). *Concepts, Kinds, and Cognitive Development,* Cambridge, MA: MIT Press.

Malt, B.C., & Johnson, E.C. (1992). Do artifact concepts have cores? *Journal of Memory and Language*, *31*, 195-217.

McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory and Cognition*, *6*, 462-472.

Medin, D.L., Goldstone, R.L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254-278.

Murphy, G.L., & Medin, D.L. (1985). The role of

theories in conceptual coherence. *Psychological Review*, *92*, 289-316.

Nosofsky, R.M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 700-708.

Rips, L.J. (1989). Similarity, typicality and categorization. In S.Vosniadou & A.Ortony (Eds.), *Similarity and Analogical Reasoning*. Cambridge: Cambridge University Press.

Rips, L.J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General*, *122*, 468-486.

Rosch, E., & Mervis, C.B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.

Sloman, S.A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology, 35*, 1-33.

Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and process in semantic Memory: A featural model for semantic decisions. *Psychological Review*, *81*, 214-241.

Smith, E.E., & Sloman, S.A. (1994). Similarity-versus rule-based categorization. *Memory and Cognition*, *22*, 377-386.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293-315.

van Mechelen, I., Hampton, J.A., Michalski, R.S., & Theuns, P. (Eds.) (1993). *Categories and Concepts: Theoretical Views and Inductive Data Analysis*. London: Academic Press.

Zadeh, L. (1965). Fuzzy sets. *Information and control*, *8*, 338-353.