SIMILARITY AND CATEGORIZATION

James A. Hampton
Department of Psychology
City University
j.a.hampton@city.ac.uk

Abstract

The importance of similarity as a basis for categorization is reviewed, and it is argued that many of our everyday concepts are built around similarity clusters.

The intuitive idea that we classify together those things that we find similar has had a chequered history in psychology. While there was considerable theoretical and empirical development of similarity-based classification models in the 1970s (Medin & Schaffer, 1978; Rosch, 1975), subsequently there has been what might be termed a "rationalist backlash" against these models. In particular a number of researchers have questioned the degree to which the notion of similarity is sufficiently clearly defined and constrained to serve as an explanation of categorization. In this paper, I discuss arguments and review evidence for and against basing categorization on a notion of similarity, and conclude that, construed broadly, similarity may be the best explanation of how most of our conceptual categories function.

Similarity-Based Categorization

What is the evidence that similarity plays a role in categorization? To answer this question 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 having three 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.

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).

There is also a feedback loop in this process. 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. It is at this point in the story that a number of psychologists have argued that something other than mere similarity and feature weights must be taken into account. 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. When medical research first tackles 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 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 biochemical malfunction -- will usually allow the syndrome to be replaced by a clearly defined disease/condition category, with its own set of diagnostic tests. 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 extended analogy with medical 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. 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 (akin to a syndrome), 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 (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 non-monotonic dissociation in the relation between similarity and categorization. If categories are formed around prototypes, then it should not be the case that one item could be more similar (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.

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 an object that was half way 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).

Rips' second example involved a creature (or artifact) which metamorphosed into something else. For example a bird-like creature was transformed into an insect-like creature through an environmental accident. When asked whether it 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.

The third example 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 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 a more extensional form of reasoning, employing intuitive statistical reasoning to find the more likely category.

There is no space in this paper to go into a detailed discussion of the validity of Rips' three cases of non-monotonicity (but see Hampton, 1997, for a fuller discussion). What is clear is that dissociations between typicality and category membership can be demonstrated albeit with relatively non-standard types of material. The first case asked people to imagine an object which is specified *only* by its size. The second involved a creature whose appearance changed, but about whose internal organs and genetic make-up subjects were told nothing, and the third case involved presenting subjects with strong cues to employ extensional reasoning using relative frequencies in their category judgments. (Physicians are familiar with the phenomenon of cases that may resemble condition A more than condition B, but where the extreme rareness of condition A means that a diagnosis of condition B is more likely to be correct.)

One aspect that all three demonstrations share is a presupposition that categorization is in fact all-or-none. Thus the object was either a coin or a pizza, it was either a bird or an insect, and either from one distribution or the other. The categorization task was always presented to the subject as one in which the *correct* categorization had to be *predicted* on the basis of the available information. As noted earlier of course, this presupposition is antithetical to the similarity-based approach where the correctness of a categorization is not something that can always be resolved. Some items are by their nature borderline to a class, and no further exploration would reveal their true nature any better.

Evidence for Similarity in Categorization

In the light of these various critiques of similarity-based categorization it is worth briefly reviewing the evidence *for* the prototype model. 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 opponents of prototype theory. 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). 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 (1996a) 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 (1997) is a further demonstration of how similarity can be shown to affect people's reasoning. In one demonstration, Sloman found that people were more likely to accept the truth of a logically necessary conclusion when the two premises were similar than when they were not. Similarity apparently pervades 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. 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.

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). Hampton (1997) 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. 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

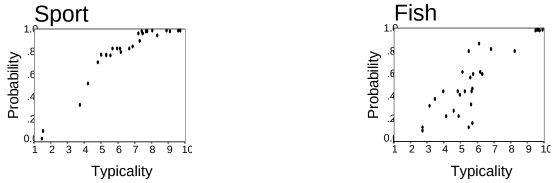

Figure 1: Illustrating the best and worst correlations between Typicality and Categorization Probability

mean rated typicality and categorization probability. Technical details can be found in Hampton (1996b). The first conclusion was that there were clear differences between individual categories in terms of how clearly categorization probability could be predicted from typicality. Figure 1 shows the best and the worst fit categories. For Sport there was a clear threshold function, with practically no systematic deviation. For Fish on the other hand, 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 in a rating questionnaire which was administered to twenty students at the University of Chicago. From this questionnaire, variables were computed for each item, corresponding to 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.

Hampton (1997) 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.

What Role Does Similarity Play?

In this paper 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 notion that our concepts should *not* be based on similarity?

There are two issues here which need to be kept separate. The first is that the world contains important distinctions that are not always immediately obvious in the outward appearance of objects. Two mushrooms may be very similar, but whereas one makes a tasty meal, the other is deadly poisonous. A crude view of similarity-based categorization would argue that we could never learn this distinction, since it would require forming a category that cuts across the way things appear to us perceptually. This view is to take *perceptual* (in fact usually *visual*) similarity as the only meaningful way of defining similarity. Perceptual similarity is indeed a very powerful and salient factor in our thinking, and it probably represents the "prototypical" or default way in which we understand similarity.

There is however a more powerful way to treat similarity, in which any dimension may enter into the computation of similarity. We might then talk of "deep similarity" as opposed to "surface similarity". If some subtle morphological characteristic of the mushrooms provided a clear predictor of the effects of eating them, then this characteristic would be given a very high weight in the computation of similarity for the purpose of culinary classification. After all there is very little similarity in the effects of eating the two mushrooms, and this factor would be sufficiently important to carry great weight in determining categorization.

The first point is therefore that similarity should be broadened to encompass a range of semantic information that goes well beyond the perceptual appearance of objects. When this is properly understood, it is clear for example why whales should not be fish. When examined more closely, when their behavior is observed and their internal organs (lungs, warm blood, brains etc.) are inspected, their similarity to other mammals, and dissimilarity from fish becomes quite obvious. There is no need for a theory of evolution to make this observation, just a curiosity about the way things are.

My second point is that as well as tending to 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 has led to the huge 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 with absolute certainty. What should be obvious to most psychologists who have attempted to study this more "advanced" type of thought 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, John Gardiner, Dedre Gentner, Doug Medin, Greg Murphy, Lance Rips, Brian Ross and Steve Sloman.

References

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.

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.

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. (1996a). Conceptual Combination: Conjunction and Negation of Natural Concepts. *Memory and Cognition*.

Hampton, J.A. (1996b) The relation between categorization and typicality: an analysis of McCloskey & Glucksberg's (1978) data. Unpublished report.

Hampton, J.A. (1997) Similarity-based Categorization and the Fuzziness of Natural Categories. *Under review*.

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.

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

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289-316.

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. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192-232.

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

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.

Sloman, S. (1997). Feature inheritance in inductive reasoning. *Manuscript under review*.

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.