Hampton, J.A. (1997). Psychological representations of concepts. Chapter in M.A.Conway (Ed.) *Cognitive models of memory*. pp. 81-110. Hove: Psychology Press.

This chapter concerns the psychological representation of the concepts that people use to classify, label and understand the surrounding world. As a first step, let us consider what we mean by concepts and conceptual representation. Shanks (this volume) differentiates categorization learning from conceptual knowledge. Categorization itself (the association of particular responses to particular classes of stimuli) is a very broad class of behaviour. At one extreme of complexity it may involve sophisticated judgments requiring lengthy argument and consideration of evidence. For example a juror in a trial must choose to categorize the accused in one of two categories - guilty or not guilty. Similarly an expert in fine art may have to categorize a painting as being a genuine work or a fake. At the other end of the complexity scale, very simple machines regularly perform categorization - a central heating system categorizes the surrounding temperature as above or below the desired level, and responds accordingly. Even simple objects and materials in the world could be said to categorize if they respond differentially to different environmental conditions - for example water categorizes temperatures into those above 0°C (where it is in liquid form) and those below 0°C (where it is solid).

We can see then that categorization per se is a behavioural measure that may tell us little about the basis on which that categorization is made. Categorization behaviour does not itself indicate any conceptual understanding. We would not wish to claim that water possesses the concept of its freezing point. Thus a system may be categorizing on the basis of complex thought processes involving deep conceptual knowledge, or it may be categorizing in a simple mechanical fashion. If we have no knowledge of the mechanism itself, then we can only characterize the classification behaviour in terms of the categorization rule that it appears to be following - that is the categories that it imposes on the stimulus input. Chater and Heyes (1994) develop this point and argue that even in the case of other species of animals, we have no way of determining what kind of concepts they possess, or indeed whether it even makes sense to speak of "concepts" in other species. With other humans, we have the additional source of information that is provided by our ability to speak to each other, and by our common assumption that our thought processes work in the same way. Thus not only can we set people categorization tasks, but we can also instruct them on the basis of categorization to use, and interrogate them on how they believe they are doing the task. We may also introspect and examine our intuitions about how we use our own concepts. We therefore have a far richer set of methods for studying human categorization. The question that has interested many psychologists for the last forty years or more is how people use concepts as the basis of categorization. The main concern of this chapter will be to consider rival views of how concepts are represented in memory and used to categorize the world.

In 1981, Smith and Medin published an account of the prevailing state of theorising about concepts in which they described three main classes of model - the Classical, the Probabilistic, and the Exemplar model (see also Hampton & Dubois, 1993). Since that time, two further related "views" of concepts have been developed -- namely the Theory-based and the Essentialist views of concepts. The following sections provide brief accounts of each of these models, and review evidence for and against each one.

A) The Classical View

The view of concepts first developed by cognitive psychologists was one that was borrowed from many sources: analytic philosophy, set logic, lexical semantics, behaviouristic psychology, and the "exact" sciences like chemistry and biology. In their ground-breaking work, Bruner, Goodnow and Austin (1956) set out to explore how people go about discovering the conceptual basis of a given way of classifying a set of stimuli. Presented with an array of stimuli which varied along a number of different dimensions, subjects in their experiments were required to select a particular example stimulus for testing. Having made a selection they would be told whether or not the stimulus fell within the category that had been defined. Subjects were also told what logical form the categorization rule would take - for example a conjunctive rule involving two "necessary" attribute values which are both required for category membership, or a disjunctive rule in which either of two attribute values would be sufficient for category membership. Bruner et al were therefore interested initially in how subjects would develop search strategies for arriving at a correct hypothesis of the rule in question.

This characterization of concepts as categorization rules was also commonly applied in behaviouristic studies conducted at the time. For example Shepard, Hovland and Jenkins (1961) devised stimulus sets based on three binary dimensions. The combination of three dimensions, each with two possible values yields 8 possible types of stimulus. Shepard et al, investigated a variety of different rules for categorizing the 8 stimuli into two classes, including simple unidimensional concepts (based on just one of the dimensions), conjunctive and disjunctive combinations of two dimensions, and more complex rules such as exclusive disjunction (either one or the other of two dimensional values, but not both). Their focus was on the difficulty of learning of these different "concepts", as subjects performed a serial task learning through trial and error to classify the stimuli in the predefined manner.

Smith and Medin (1981) characterized the Classical View of concepts as one in which it is assumed that conceptual categorization is based on logical classification rules. More specifically this view of concepts typically assumes that the logical structure of most natural concepts is a conjunction of necessary features.

The notion of "feature" had been borrowed from its use in structural linguistics. Early theories of lexical semantics such as Katz and Fodor (1963) had proposed that the meaning of nouns could be characterized in terms of a limited set of defining features. Thus the meaning of the word BACHELOR could be represented by the set of features:

HUMAN +
ADULT +
MALE +
MARRIED –

Note that in this treatment, a semantic feature is normally a dimension with just two values - present (+) or absent (–).

The Classical View of concepts thus presented a very "clean" view of semantic representation (and hence of conceptual structure). Conceptual categories are defined in terms of the presence or absence of a conjunction of defining features. Thus all category members have the same set of defining features in common, and any instance that lacks any of the defining features is by definition not in the category. The advantages of this type of conceptual structure are readily apparent in the exact sciences. For example the classical approach to classification of the variety of biological kinds, was to divide the natural world into major divisions, which were each then divided into smaller, mutually exclusive, subclasses in what is known as a taxonomy. The principle of a classical taxonomy is that each class "inherits" all the defining features of the classes above it - the superordinate classes. Thus a SPANIEL will have all the defining features of DOGS, all the defining features of MAMMALS, and all the defining features of all the remaining superordinate classes of ANIMAL, and BIOLOGICAL KINDS. As a model of concepts this taxonomic structure has many advantages - it is a highly efficient system for storage of information, as featural information need only be represented at one level of the taxonomic hierarchy, rather than for all the different classes of which it is true. It also permits rapid semantic decisions to be made simply by reading them off the taxonomic tree. Thus the following kinds of sentences could all be easily verified by a straightforward "look-up" procedure:

All spaniels are mammals
All cats have kidneys
No cats are dogs
No cats have gills

The classical view of concepts together with the assumption of a taxonomic storage structure was used by Collins and Quillian (1969, 1972) to develop an early model of "semantic memory" (a term introduced by Tulving, 1972 to refer to that part of our memory which stores knowledge of the world and the meaning of words, as opposed to memories of particular episodes or events). They represented the taxonomy as a "semantic network", involving concepts as nodes, connected by directional links indicating two kinds of semantic relation - "is a type of" and "is a property of". Collins and Quillian showed that the time taken to verify sentences involving superordinate classification (All cats are animals) or property possession (All cats breathe) reflected the distance in the taxonomic structure between the head noun term (cats) and the level at which the information was stored. The advantages of this form of structure were obvious and similar structures are commonly found in artificial intelligence systems for knowledge representation.

Evidence and critique

Beyond its intuitive appeal, there is not a lot of evidence directly supporting the classical view -- probably because many researchers took it to be self evident. For those who still support the view (Armstrong, Gleitman, & Gleitman, 1983, Osherson & Smith, 1981, 1982), the main arguments in favour of it are arguments against the other views - in particular the probabilistic view described below (see also Fodor, 1994). There are however serious difficulties with maintaining the classical view.

First, it is fair to say that the classical view was heavily dependent on biological kinds as a model for all concepts. However, when applied to the full range of concepts that need to be represented, the suitability of a taxonomic structure becomes highly suspect. How do we categorize the range of objects found in a domestic household for example? There are many cross-classifications that would be possible - in terms of their size and weight, their function, their value, their material. To select any one feature and place it at the head of a taxonomy would be an arbitrary choice. Even the hierarchies explored by Collins and Quillian (1969, 1972) were shown to be problematic. Collins and Quillian took the structure of the taxonomy to be self evident, and set about exploring times taken to compare concepts at different levels. However when the problem was approached more systematically it was found that categorization times were not always predictable from distance in the taxonomy. For example Smith, Shoben and Rips (1974) showed that in some selected cases, a more distant category relation could be verified faster than a closer one. Subjects were quicker for instance to judge that "A chicken is an animal" is true, than that "A chicken is a bird" is true. It was not therefore possible that subjects were inferring the former by retrieving the two facts "A chicken is a bird" and "A bird is an animal", and then performing a transitive inference. Conrad (1972) found that the associative frequency of properties to nouns was a better predictor of verification time for properties than was distance in the taxonomic tree. Hampton (1982) and Randall (1976) found evidence that even the common assumption that taxonomies represent class inclusion hierarchies could be challenged. Hampton's subjects agreed that CHAIRS are a type of FURNITURE, and that CARSEATS are a type of CHAIR, but were then unwilling to allow that CARSEATS were a type of FURNITURE. It follows that if there is a taxonomy with classes ordered FURNITURE - CHAIR - CARSEAT, there can not be complete inheritance of the defining features of superordinate classes by those below.

The second source of difficulties was the lack of easily available defining features for many common concepts. Hampton (1979) asked subjects to provide feature information for eight common categories like FRUIT, FISH or FURNITURE. They were interviewed individually and asked to think of features that were important to the definition, features that might make some examples more or less typical, and features that might be relevant at the borderline of the category in determining whether some instance was in or out of the category. From the resulting feature lists, Hampton drew up an inclusive list of features for each category and had another group of subjects judge the extent to which a set of potential category instances possessed each feature. Finally a third set of judges made categorization judgments about the same set of instances. It was clear when comparing the three sets of judgments that for most of the categories there was no conjunctive combination of defining features which would successfully separate category members from category non-members. Most of the features that subjects generated were of one of two kinds -- they were either very general features such as "Used by people" or "Living", which were too general to pick out the category from others in the same domain, or they were features such as "Tastes sweet" (of FRUIT) or "Flies" (of BIRD) that were typically true of many category members but not of all. It therefore appears to be the case that people's representation of featural (intensional) information does not contain features which could provide a classical common element definition, but does contain a lot of other information about what kinds of properties are commonly found across category members. It was as an answer to this kind of problem with the classical view that the probabilistic view gained credence.

B) The Probabilistic View

The probabilistic view of concepts arose largely from difficulties in applying the classical view to the actual concepts that people possess, as opposed to those which they would possess in some ideal world. Once again, the roots of the view can be traced to a number of other disciplines. In philosophy, Ryle (1951) and Wittgenstein (1958) discovered that many interesting concepts -- like work, or game, or individual identity -- were frustratingly difficult to pin down within the classical system of giving definitions. Whenever one proposed an apparently defining feature of such a concept, counter-examples would immediately spring to mind. The reader may like to consider how one would define the essential nature of an activity which would make it count as WORK. Perhaps work always involves some effort - but then a night watchman is working even when simply sitting in a chair drinking coffee. Perhaps payment for one's time may be a required feature, but voluntary and charity workers are presumably still engaged in work. Interestingly, similar problems had arisen within biological classification, where the standard taxonomy developed by Linnaeus was found to create all kinds of anomalies. Disputes arose between different types of taxonomy, and a new form of taxonomy was developed -- numerical taxonomy -- which used the statistical properties of similarities amongst organisms to derive classes, rather than selecting the major divisions amongst classes a priori (Jardine & Sibson, 1971).

In cognitive psychology, the probabilistic view of concepts was developed largely through the work of Eleanor Rosch and Carolyn Mervis, who proposed the Prototype Theory of concepts. Rosch (writing initially under the name of Heider) had spent some time studying the basis of colour and shape concepts in a primitive culture - the Dani in New Guinea (Heider, 1972). She had observed that colour concepts like RED and BLUE appeared to be natural "anchor points" or prototypes on the colour spectrum, such that even although the Dani had only two colour words in their language, and no words corresponding to RED or BLUE, they found it much easier to learn to categorize colours around these anchor points than around intermediate colour hues (for which few if any languages in the world have specific colour terms). Rosch and Mervis (1975) went on to develop the notion of prototype to apply also to common noun categories like BIRD, FRUIT or VEHICLE. These were just those categories which philosophers like Wittgenstein had proposed as examples of concepts with no clear defining features. Wittgenstein (1958) had pointed out that although it was difficult to define the boundary of such categories (that is to find an exact rule for classification), no-one would dispute the classification of clear typical examples. He had suggested that such categories were similar in structure to the way that people in the same family resembled each other -- there might be a characteristic family nose, and family eye colour, and family hair type, and different family members would be expected all to have at least some of these characteristics, although not all.

The Prototype theory of concepts as developed by Rosch and Mervis (1975) therefore proposed that categorization in many everyday concept categories might follow similar "family resemblance" principles. A particular type of object might be classified as a FRUIT or as a VEHICLE on the basis of a variety of different characteristics (or features), none of which alone need be common to all category members, or sufficient for categorization. In terms of the formal or logical structure of such categories, the prototype view can be seen as involving a set of category features and a categorization rule that states that an instance belongs in the category provided that it possesses a sufficient number of those features. It is therefore still a logical classification rule which can be stated quite explicitly, although it would be inelegant to express it purely in terms of a conjunction and disjunction of features.

An alternative way in which prototype categories have been defined is in terms of similarity. Similarity is commonly assumed to depend on the proportion of common versus distinctive features between two objects (see for example, Tversky, 1977). Thus the prototype representation of a concept could be seen also as an "idealized" instance, together with a way of measuring the similarity of other instances to that ideal prototype (that is to say a "similarity metric"). The categorization rule is then specified in terms of whether similarity of an instance to the prototype for the category exceeds some criterion value. While the family resemblance characterisation of a prototype as a set of attribute features can be translated into the similarity based characterisation, the reverse is not always the case. For example colour terms like RED are presumed to depend on similarity as measured in terms of psychological distance along the hue spectrum. Other stimuli have also been created for use in psychology experiments that are not easily translated into a featural description. For example Posner and Keele (1968) used a set of dots randomly distributed around a matrix as the "prototype" stimulus, and then defined the categories of stimuli to be learned in terms of the amount of distortion introduced to this random prototype by moving individual dots to neighbouring positions in the matrix. Shanks (this volume) gives more details of this type of prototype stimulus structure. For this kind of stimulus, a prototype concept can be represented as a region within a similarity space, organized around a central "prototype" point, but cannot easily be described in terms of family resemblance.

The "family resemblance" notion of prototype and the "region of similarity space" notion of prototype concepts are not equivalent, and care needs to be exercised not to confuse the two. Osherson and Smith (1981, 1982) provide a formal account of prototypes in which the similarity space figures as a part of the definition, whereas Hampton (1993) prefers to specify prototypes as attribute descriptions. Given that many of the concepts that we are chiefly interested in are complex semantic entities rather than simple unanalysed visual shapes, the attribute description approach is likely to provide a richer characterization of the information available in conceptual memory. Similarity space representations place severe constraints on what is represented (the prototype is a vector of coordinates on a set of dimensions). There are also many difficulties to be found in mapping similarity directly into metric spaces (Tversky & Gati, 1982).

Prototype theory is thus a different way of providing a classification rule for categorization -- one that is based on a cluster of attributes, or on a similarity criterion. The advantages of this approach are two-fold. First, it fits better with what is known about the kinds of property information that people generate when asked to define concepts. As discussed above, when defining concepts, people do not distinguish between properties that are common to all members of a category and those that are just common to many members. In one unpublished study, Barsalou (personal communication) asked one group of subjects to generate only the most defining attributes of a class, while a second group generated any attributes that they felt were associated with it, however loosely. Barsalou found no reliable difference between the two sets of attributes generated. Second, it provides a ready explanation of why there are many <u>borderline</u> cases of classification. In any similarity based categorization, there is always the possibility that a particular instance will have a similarity to the prototype (or a number of matching features) which is very close to the criterion for category membership. In such a case then small variations in individuals' concept representations or small contextual effects on the weight of different features or dimensions, or on the membership criterion itself, will result in <u>instability</u> in categorization. Subjects will feel unsure whether the instance really should be in the category or not. There is plenty of evidence that this kind of borderline instability occurs. McCloskey and Glucksberg (1978) asked a group of subjects to categorize lists of category instances in 18 different categories on two occasions. They found that for items that were judged to be good members of the categories, subjects were quite consistent, but that for items that were atypical, categorization was probabilistic -- subjects disagreed with each other, and often changed their decision from one occasion to the next.

Rosch and Mervis also introduced two important new notions in their theory. The first notion has already been referred to, and is the idea that category members differ in their <u>typicality</u> as members of the category. Rosch (1975) asked subjects to consider a list of instances of a category and to judge the "goodness of example" or typicality of each instance as an example of the category. These typicality ratings were made on a seven point scale, with 1 as "Highly typical" and 7 as "Not a member of the category". The mean typicality ratings were shown to be highly reliable (although as Barsalou, 1987, later pointed out, reliability of the mean ratings was partly just a function of the sample size from which the means were calculated). Furthermore, compared to atypical items, typical category members were shown to have a higher degree of "family resemblance" to other category members and a lower degree of family resemblance to instances from contrasting categories (Rosch & Mervis, 1975). Further research has established category typicality as a major dimension of semantic memory, influencing a wide range of experimental measures and tasks (for example: categorization time, inductive inference, anaphoric reference, and episodic memory).

The second notion introduced by Rosch and Mervis was the idea of a <u>basic level</u> of categorization. In a variety of different tasks, Rosch, Mervis, et al. (1976) were able to show that a particular object was most easily categorized at an intermediate level of categorization that they termed <u>basic</u>. Thus a particular object might be most naturally or readily categorized as a CAR, and less easily categorized as a VEHICLE (superordinate level) or as a VOLKSWAGEN GOLF (subordinate level). Compared to higher or lower levels, basic level categories were shown to have higher frequency single word names, to be the terms most often used by parents in naming objects for children, to be the most general level at which a generic image of the object class could be formed, and to be the level at which picture-word matching was most rapid.

Subsequent research has suggested that basic level terms are also the most general categories where the kinds and arrangement of parts are the same (Tversky & Hemenway, 1984). It has also been shown that the notion of basicness may not be correctly applied to a whole level of categorization, since the term that is basic for an object may vary as a function of typicality, as defined above. For example, when considering SPARROWS, THRUSHES and BLACKBIRDS, Rosch's results suggest that "Bird" is the basic level term. It has a higher word frequency, and provided the quickest word-picture matching. However if atypical birds are considered such as OSTRICH or PENGUIN, then it is highly likely that the basic level term is no longer "Bird", but is "Ostrich" or "Penguin" (Hoffman, 1982, Jolicoeur et al., 1984). Some caution should therefore be exercised in using the notion of Basic Level as if it applied generally to all the terms at a particular level of a taxonomic structure.

Evidence and critique

Much of the evidence for the prototype theory of concepts derives from an extensive series of experiments performed by Rosch, and by Rosch and Mervis, and published in 1975 and 1976. They established the importance of typicality as a psychological dimension of natural categories, and the relation of typicality to family resemblance. Hampton (1981) investigated eight abstract categories, like RULE, INSTINCT and SCIENCE, and found that some concepts appeared to have family resemblance structure whereas others did not. Other research showed that prototype concepts were used in a variety of other domains - for example, personality perception (Cantor & Mischel, 1977, 1979; Hampson, 1982), categorization of everyday situations (Cantor, Mischel, & Schwartz, 1982), and psychiatric diagnosis (Cantor, Smith, French, & Mezzich, 1980).

Criticism of prototype theory and the probabilistic view has come from three main sources. First there is the problem that the simple feature list representations assumed by prototype models are clearly inadequate as full representations of conceptual knowledge. We know and understand a great deal more about a biological kind such as BIRD than a simple list of descriptions about its common appearance, behaviour and habitat. For example we know that it has wings and that it flies, but also that it uses the wings in order to fly. (In fact the reason that we call its upper limbs "wings" has precisely to do with this functional relationship). As with manufactured objects, we have an understanding about structure-function relationships that involves theoretical notions of cause, and purpose. Our concepts are thus embedded in deeper theoretical domain knowledge (see the Theory-based view below). One way to develop this idea is to use a more sophisticated form of knowledge representation - such as the notion of a frame taken from Artificial Intelligence. Barsalou (Barsalou, 1992; Barsalou & Hale, 1993) developed the argument for a complex system of frame-based representation of concepts, in order to account for our ability to use concepts in planning and achieving goals. Note that more powerful representational forms are not necessarily incompatible with a similarity based categorization rule involving a "prototype". The central notion of a prototype need not be tied to any one particular form of knowledge representation. A second criticism of feature lists is that they are not only too simplistic, but also appear to involve a degree of circularity - one is "explaining" categorization with respect to one category (Bird) by appeal to a series of other categories (flying, feathers, head, legs, eggs etc.) many of which probably depend on the category itself for their definition. Birds have heads, but not just any head will do. Whether or not this move is getting us any nearer to our goal of modelling conceptual representation is debatable (Fodor et al., 1980). Future development of the theory will need to move away from purely linguistic descriptions of "features" and replace them with semantic values that are better grounded in a sub-symbolic level. Just as a foreigner could never learn a language from a dictionary since the meaning of each word is given using other words in the same dictionary, so the features themselves need to be translatable into non-linguistic representations related to action and perception, if the representational system is to avoid circularity. Barsalou (1993) has developed a proposal that visual imagery may be able to provide such a pre-linguistic representational language for concept representation.

The second critique of prototype theory and the probabilistic approach to concepts comes from a seminal article by Osherson and Smith (1981) which set in train a new research program into the problem of conceptual combination. Osherson and Smith investigated the claim that probabilistic concepts could be treated formally as fuzzy sets, using the axioms of fuzzy logic developed by Zadeh (1965). Fuzzy logic is a system of logic in which membership of a set, and the truth of a statement can take continuous values intermediate between true (1) and false (0). An example is the truth of the statement "John is tall". In classical logic, this statement must be either true or false (or meaningless). However this would require the fixing of a particular value for height such that everyone above some height is tall, and everyone below that height is not tall. While possibly making life easier for logicians, this kind of fixed criterion does not capture the way that we normally understand the truth of such statements. It is more the case that membership in the set of "tall men" is graded, and increases from a clear false for a 5-foot-tall fully grown individual, to a clear true for a 7-foot-tall basketball player. Zadeh proposed that such intuitions could be captured by assigning a value on the range from zero to one to represent "how true" the statement was. Osherson and Smith (1981) picked up on various suggestions in the prototype literature that membership of natural categories like FRUIT or FISH might also be modelled by fuzzy logic. In particular it had been suggested that typicality gradients and borderline uncertainty were the kind of phenomena that lent themselves well to a fuzzy characterization.

Osherson and Smith considered a number of the axioms of fuzzy logic as applied to natural concepts. The best of their arguments concerned the way in which set conjunction could be applied to prototype concepts. The conjunction of two sets is the class of instances that belong in both sets - thus the conjunction of Pets and Fish is the class of Pet Fish. Without going into the technical details of the argument, fuzzy logic proposes that the degree of membership in a conjunction must always be less than or equal to the degree of membership in either constituent set. It should, after all, be at least as difficult to achieve membership in two sets at once as it is to achieve membership in just one of those sets alone. However people's intuitions of typicality do not fit this stipulation. Common pet fish like a GUPPY or a GOLDFISH are considered to be more typical examples of the conjunction PET FISH than they are either of PETS (cats and dogs being the most typical) or of FISH (salmon and cod are more typical). On the basis of this and other examples Osherson and Smith argued that the probabilistic approach was not amenable to a logical treatment, and could therefore not be taken seriously as a basis for conceptual thinking. Set logic, they contended, requires that concepts have some kind of classical logical definition as their "core". (Fodor, 1994, makes similar claims about the failings of prototypes to provide us with a sufficient basis for a compositional semantics.)

If we are to return to the classical model of concept definitions for the purposes of conceptual thinking, then what are we to make of all the evidence for non-defining properties being represented, and variations in typicality of instances? The answer proposed is that concepts have not only a classically defined core, but also a set of "stereotypical" information which is used to drive typicality judgments. Hampton (1988, 1995) describes this as the "binary view" of concepts. The view makes a distinction between issues of category <u>membership</u> -- what is and what is not an example of the category -- and category <u>typicality</u> -- how similar something is to a prototypical or stereotypical example. A version of this distinction was originally proposed by Smith, Shoben and Rips (1974), and it has been supported by several others (Armstrong et al, 1983; Landau, 1982; Miller and Johnson-Laird, 1976; Rey, 1983, 1985). It is also suggested that the category stereotype may be of particular value in the case of real-life situated categorization. We are often unable to tell the "true" nature of something by its external appearance, but the stereotype information allows us to make a good probabilistic guess as to what kind of thing we are dealing with, and hence how to react to it appropriately. Hence the stereotype has sometimes been termed the "identification procedure" for the concept.

The binary view tries to have the best of both models - it can use core definitions to ensure proper set logical relations for conceptual thinking, and it can use the stereotypical information to account for the wide range of typicality effects. It still however suffers from the central problem of the classical model - the difficulty of identifying core definitions for more than a handful of concepts. This has led some to postulate that people's representation of these core definitions may actually be empty (see the Essentialist view described below). There is also some evidence against the binary view from the literature on conceptual combination (see Hampton, 1996; Rips, 1995, for reviews). Hampton (1987, 1988) provided evidence that the way people categorize items in conjunctive concepts (like A Sport that is also a Game) was not in fact constrained by traditional set logic. People were frequently willing to accord membership in the conjunction to items that they failed to classify as belonging to one of the two sets. Thus CHESS was not considered to be a SPORT, but <u>was</u> considered to be a SPORT WHICH IS A GAME. Membership of conjunctions it appears is also flexible and subject to similarity comparisons - a conclusion that led Hampton (1988) to propose that conjunctions are understood by the formation of a Composite Prototype constructed from the two prototype representations of the individual constituent concepts.

The third critique of prototype theory came from a paper by Armstrong, Gleitman and Gleitman, (1983) in which they demonstrated that typicality gradients were not specifically restricted to prototype concepts. They asked subjects to judge the relative typicality of the members of well-defined categories, like odd or even numbers. They argued that if typicality reflects similarity-based categorization, then where categorization was clearly not based on similarity, there should be no reliable typicality effects. In contrast, they found that typicality differences for odd numbers were just as strong and as reliable as they were for "Roschian" categories like FRUIT or FURNITURE. Hence, they argued, typicality may not be taken as evidence for prototypes.

Armstrong et al.'s critique makes an important point. Typicality gradedness alone is not sufficient evidence for inferring a prototype structure. However it also fails to prove that in the case of Roschian concepts typicality may not <u>also</u> reflect similarity-based categorization. Barsalou (1985) investigated the determinants of typicality in both natural categories and what he called "ad hoc" categories (categories like "Birthday presents") using a regression approach. He found that the representativeness of exemplars was of prime importance in predicting typicality ratings for natural categories, but that some other variables were also significantly involved - for example the frequency with which someone encountered a particular example. In the case of well defined categories like odd number, it will be the case of course that "representativeness" will be constant - all odd numbers are equally good at fitting the rule for determining category membership. In this case, then other dimensions will take over the determination of typicality, with more frequently used numbers like 3 or 7 being considered more typical than unusual numbers like 831.

<u>C) The Exemplar View</u>

The third of the approaches to concept representation is the Exemplar View. Described simply, exemplar models propose that conceptual categories are represented by collections of individual exemplar representations. A standard theory of exemplar representation was proposed by Medin and Shaffer (1978), and extended in Nosofsky's Generalised Context Model (1988). Other proponents of the view include Brooks (1978, 1987) who argued for a Non-analytic approach to conceptual structure, meaning that much of the time conceptual tasks depend on the use of actual remembered individual instances situated in a particular context. For example if you were faced with some novel type of exotic fruit, your ability to classify it as a fruit might well depend on processes of comparison of the object along different dimensions of similarity with remembered examples of other common fruits with which you are familiar. According to this view, categorisation would not depend on an explicit logical classification rule (as in Classical theory) nor on comparison to any generic category representation (as in Prototype theory) but rather on a series of comparisons to particular remembered individual exemplars. The classification rule would therefore be stated in terms of the overall similarity of a stimulus to stored exemplars of the category compared to the similarity of the exemplar to stored exemplars of alternative categories.

It is clear that this proposal can readily explain the same evidence that led to the rise of the probabilistic view. When asked to generate concept definitions, subjects would in fact retrieve memories of individual exemplars and offer descriptive properties which they find tend to occur reasonably often across exemplars. The difficulty that subjects have in framing successful definitions is entirely in keeping with an exemplar representation. Similarly the existence of borderline cases is easily explained within the exemplar approach. Some object may turn out to share similarities with the members of two distinct categories, and hence its classification in one or the other may be unstable and subject to individual and contextual variation.

A demonstration of exemplar-based effects comes from a study by Medin and Shoben (1988) in which they asked subjects to judge the typicality of different objects in subclasses of a category. For example, subjects were asked to rate the typicality of a variety of spoons in the two classes SPOON or LARGE SPOON. Medin and Shoben showed that typicality in the two classes varied as a function of another unrelated dimension - namely what the spoon was made from. They showed that while metal spoons were more typical of SPOON, wooden spoons were more typical of LARGE SPOON. The most likely explanation of this effect is that subjects are retrieving familiar instances of large and small spoons and noting the material from which they are made.

Evidence for the exemplar view is reviewed in Shanks' chapter in this volume. Much of the evidence relates to classification learning tasks where subjects learn to classify a set of well controlled, artificially constructed stimuli over many trials. Such experiments frequently show exemplar-based effects, such that similarity to a particular learned exemplar may have a more powerful influence on the speed and probability of a positive classification, than does similarity to the central tendency or prototype of the stimulus set. Exemplar models provide a good fit to many results in the classification learning area. For example, Medin and Schwanenflugel (1981) investigated the learning of two stimulus sets. One could be learned with a prototype-based rule, since the categorization was "linearly discriminable" -- a classification rule could be given based on the sum of matching features. The other task could not be learned through forming a prototype, as the categorization structure was not linearly discriminable. (To achieve this, a pair of features would contribute to an item's being in category A if they were both present or both absent, but to category B if either one were present alone). The results showed no advantage for the linearly discriminable stimulus set, from which Medin and Schwanenflugel concluded that subjects were not forming generic prototype representations in this learning task.

Another interesting result was reported by Barsalou, Huttenlocher and Lamberts (1995). Their subjects viewed a number of exemplars of tropical fish displayed on a computer screen, and learned about the different individuals. They were tested in a transfer task to judge which of various novel exemplars were more likely to come from the same category as those seen previously. The stimuli in the training set involved four very similar fish, and one very atypical looking one. When the five different fish were presented equally often, (and when they were easily differentiated), then transfer to new stimuli was driven by the more common pattern of the four similar fish. However when the atypical fish was presented more frequently than the rest, transfer was affected more by similarity to the frequent atypical exemplar. It therefore appears that subjects are affected by the storage of individual traces, as an exemplar model might predict. In one experiment, however, Barsalou et al varied the instructions given to subjects. One group were told that all stimuli seen were <u>different individuals</u>, so that the frequency of occurrence reflected the actual frequency of those stimuli in the category at large. The other group were told that there were only five individuals which were seen repeatedly, and that one (the atypical looking one) was less shy of people, and so was more likely to be observed. The results showed a strong effect of this manipulation of instructions -- the first group transfered on the basis of the frequency of encounter of the stimuli, whereas the second group were able to discount frequency and base their transfer on the fact that of the five individuals, four were similar to a prototype and one was not. (This result was only obtained under quite particular conditions -- notably when the four similar individuals were identical to each other and differed from the atypical individual on eight different dimensions, and when two additional dimensions provided clear individuating information to allow all four similar fish to be clearly differentiated from each other.) The Barsalou et al. study highlights a potential ambiguity in many exemplar models, in how they define what is to count as an exemplar. It appears that under certain conditions, subjects may be able to choose either to treat each occurrence of an exemplar as an independent exemplar trace, or to treat occurrences of the same individual as contributing to a single exemplar representation.

Exemplar models lend themselves to modelling with neural network models of classification learning. Neural network models in fact tend to blur the distinction between exemplar and prototype representations (see also Barsalou, 1990, on the points of similarity and difference between these models). In a simple neural network set up to learn a classification of stimuli, the common structure is for an input layer of nodes in which the full characterisation of the stimulus is represented to be connected to an output layer with just two nodes (Category A versus Category B) via a "hidden layer" of nodes in which the learning itself occurs. Depending on the number of nodes allowed in the hidden layer, then the network will show either more or less exemplar learning. With a large number of hidden layer nodes, then the network can learn what is effectively a "look-up" table of the classification of each individual stimulus - thus showing completely exemplar-based categorization. With a much smaller number of hidden layer nodes, the network is forced to find general patterns in the stimulus array, and to respond to similarities amongst the stimuli in each set, and the differences between stimuli in the different sets. In fact when the hidden layer has just one node, the network is only able to learn to discriminate stimuli on the basis of a weighted average of the activation from the input nodes (a linearly discriminable classification) -- which is directly equivalent to a simple prototype model.

Evidence and Critique

The topic for the current chapter is the representation of conceptual knowledge rather than classification learning per se, and so one has to question the value of exemplar models for the representation of naturally occuring concepts and instances. Evidence from classification learning is not always directly relevant to the question of natural concepts. For example, as described previously Medin and Schwanenflugel (1981) showed that linearly discriminable sets are no easier to learn that non-linearly discriminable sets. But this doesn't address the question of whether natural concept classes are linearly discriminable in terms of their attributes, and hence whether prototype models are adequate representations of natural concept classes. The learning of new categories through a training procedure has little relation to the evolution of natural concepts during the cultural and linguistic development of a society.

In support of exemplar representations for non-artificial concepts, Brooks (1987) has shown that in the classification of actual objects, there is a processing advantage for objects that have been seen previously over those that are novel, even when the previously categorised objects are wildly atypical. Thus it appears that people can remember the categorization of individuals, and do not need to reconstruct the category decision each time. However as a general model of concepts, exemplar theory appears to leave too much out. It offers no account, for example, of how conceptual categories are formed in the first place. This is also a problem for many other models (see the Theory-based account below), but the exemplar model has no account of why people with differing day-to-day experiences do not end up with widely different concepts, or why we end up with categorized experiences at all. The model also needs to assume a similarity metric (otherwise new instances could not be compared with old), which presupposes an underlying semantic structure which is not made explicit. The similarity amongst experimenter constructed stimuli may be relatively easy to control and its basis understood, but when it comes to categorization of objects in natural categories of biological kinds, or artifacts, or of events or actions in other more general categories, the basis on which similarity to stored instances is judged is far from clear. If this computation of similarity is based on underlying attribute representations of the instances (of the kind used in family resemblance prototypes) then it would appear that the exemplar view needs to specify just what that information is and how it is used to compute similarity. If the computation is not based on attributes, then it remains mysterious how unanalysed stimuli can be compared in order to compute a sensitive measure of similarity.

Exemplar models are also frequently vague about just what an exemplar is. There is an ambiguity between types and tokens of exemplars which is not always resolved (see for example Barsalou et al, 1994). When storing exemplars of spoons to represent the conceptual category of SPOON, do I store every occasion on which I see any individual spoon as a separate "exemplar", or do I keep a track of different individuals, but represent each individual only once? Alternatively I might keep track of different indistinguishable subtypes of exemplars - there are, after all, usually many spoons in the kitchen drawer that all look alike. Many exemplar based effects in natural categories (as opposed to classification learning tasks) may in fact reflect the learning of sub-categories of concepts -- for example that the category of spoons can be divided into teaspoons, soup spoons, cooking spoons and so forth -- rather than a true exemplar representation. Barsalou et al.'s (1994) results suggest that in learning a new category people are heavily influenced by the availability of exemplars in memory (an example of the availability heuristic identified by Tversky & Kahneman, 1973). It remains to be seen to what extent conceptual categories in long term semantic memory are also influenced by such factors.

D) The Theory-based and Essentialist views

The last two approaches to conceptual structure will be described together. Although there are different points of view within this approach, they share the major claim that the other views of concepts are too simplistic and reductionist in their approach. The theory-based approach was introduced by Murphy and Medin in a seminal paper in 1985. They posed the important question - why do we have the concepts that we do, and what makes them coherent? The answer, they claimed, was that concepts are not simply arbitrary ways of dividing up the world into similarity-based classes for the convenience of giving things names. Rather, concepts provide a vital function in helping us to understand the world around us. By understanding, we mean that concepts play a central role in allowing us to construct theories of the world. The success of any theory is dependent on the appropriateness of the concepts from which it is constructed. Hence, Murphy and Medin argued, we have developed particular ways of classifying and labelling our world in order to maximize our ability to understand and explain it. Rips (1995) characterizes this approach in terms of "explanation-based" classification. He proposes that we classify an object into a conceptual category that is best able to explain why the object has the attributes that it does. Concepts come with "mini-theories" that are able to provide explanations of the set of properties displayed by an instance. Whichever concept provides the best explanation is the one that the instance is most likely to belong to.

Note that the Theory-based view of concepts denies the importance of similarity as an organizing principle in forming categories. There are many classes of object and event which we class together on the basis of deeper aspects, regardless of superficial similarity. This theme has been developed particularly strongly in work with adults by Rips (1989) and in the developmental literature on children's acquisition of concepts (Carey, 1985; Gelman, 1988; Keil, 1989). Rips (1989) developed a number of arguments against similarity as a basis for conceptual categorization. In one set of studies he showed a dissociation between judgments of similarity and judgments of category membership. For example, Rips asked his subjects to tell him in inches the smallest diameter that they thought a pizza could be, and the largest diameter that they thought a quarter (an American coin) could be. Subjects were then asked to imagine an object half way between these two sizes, and asked (a) to judge its similarity to a pizza or a quarter, and (b) to say which it was more likely to be -- a pizza or a quarter. Similarity judgments did not follow category membership, in as much as subjects used the inherent variability of the categories in judging membership (pizzas being more variable than quarters), but subjects ignored variability in judging similarity. (Smith and Sloman, 1994, have since replicated this study and find that the range of situations in which the dissociation occurs is relatively restricted -- in many situations subjects continued to categorize by similarity).

In another study, Rips (1989) told subjects a story about a creature called a Sorp which looked and acted pretty much like a bird, but which lived on a radioactive waste dump. Over the years, the sorp suffered a metamorphosis, losing its feathers, and growing four extra legs until it had an insect-like appearance. Although subjects were happy to say that the sorp was now more similar to an insect than to a bird, they were more likely to categorize it as a bird. These studies make the point that in classification people pay attention to more information than simply the surface appearance of the object being categorized. Its history and ontogeny may be more influential than its current state in determining how it will be classified.

In the developmental field, there have been a number of similar demonstrations of the importance of "deeper" more theoretical knowledge in determining how even quite young children choose to classify. For example, Carey (1985) showed children aged between 4 and 10 a mechanical toy monkey that could move its arms to bang a pair of cymbals. Although they judged the monkey to be more similar to people than any of a range of other animals, all except one group of 4 year olds nonetheless denied that the toy could breathe, eat or have babies. Gelman and Markman (1986) similarly showed that when told some new fact about a reference object, four year olds were more likely to generalize the fact to an object from the same general category as the reference object, than to another object that was superficially more similar to the reference. Their willingness to induce properties was driven by categorization rather than similarity. Hence it can be argued that categorization cannot be simply based on similarity.

Keil (1986, 1989) showed children a series of pictures of a horse which is painted with stripes until it ends up looking like a zebra. Children then had to say whether it was now a horse or a zebra. While kindergarten children went along with the transformation and accepted that the horse had turned into a zebra, by the age of 8 years most children believed that this kind of transformation would not change what kind of animal it was. In contrast, when an artifact kind, like a coffee pot was similarly transformed until it looked like a bird feeder, the older children (and adults) accepted that the object was no longer a coffee pot. It thus appears that we may have different kinds of definition for natural biological kinds and for artifact kinds, and that our concepts for objects in each domain rely on different underlying theories about what makes something the kind of thing that it is. Keil (1981, 1989) has speculated that there may be a small number of such domain specific theories, which are available from a very early age and help to determine the path of conceptual development. He argues that, in the same way that Chomsky (1980) has argued for an innate basis for the learning of the deep aspects of linguistic structure, so there must be an innate structure to our conceptual understanding in order to get the inductive process started. The statistically-based clustering of "similar" objects, (what Rosch called the "correlational structure of the world"), as exemplified by neural network classification learning models, can only begin to work once a clear domain of input has been selected. The input vector to the network must be of restricted size, and must contain information that is coherent and relevant to the classification of the domain. There is currently no satisfactory proposal as to how such organisation could arise through the process of learning alone.

The final view of concepts to be described in this brief overview is the Essentialist view. Discussion of the nature of concepts has a long tradition within philosophy, and a rival view of concepts has been developed on the basis of insights arrived at by Putnam (1975) and Kripke (1972), which takes a more radical approach to the rejection of similarity as a basis of concepts. The Theory-based view rejects appearance and surface features as being of incidental relevance to the "true nature" of things and their categorization. Instead it proposes that categorization is based on less obvious aspects of the objects. In the case of biological kinds this might be the anatomy and physiology of the organism, while in the case of artifacts it could be the purpose and function for which the object was originally created, and is currently used. According to the Essentialist position, even these deeper, more theoretically relevant aspects still do not capture the real definition of a concept. Consider a cat. One of the features in common to all cats may be that they have a liver. But suppose that a mutant creature appeared which was like all other cats in all respects except that the physiological functions normally performed by the liver were somehow performed by other organs in the body -- it had no liver. It does not appear likely that we would automatically wish to reject the creature as being a cat. But the same argument could be applied to any other deep "innards" based aspect of an instance of cat - and so one may not presume that the theory-based definition of the concept is truly capturing the definition of the kind any more than the surface features of fur, ears and purring do.

The doctrine of Psychological Essentialism as proposed by Medin and Ortony (1989) argues that people have essentialist beliefs for many concepts. That is to say that they believe that both the internal and external properties that we observe in a class of things are present as the result of some common hidden essence, which defines the true nature of the class. A cat is a cat, according to this view, because all cats have some essence in common -- but that essence is itself not directly identifiable with any particular set of observable properties. Rather the essence reveals itself through the occurrence of common regularities and similarities amongst the members of the concept category (note the similarity with Rips' account of finding the concept that best explains the instance). Medin and Ortony argue that we understand that there are these essences, even though we may be quite ignorant as to what they are in any particular case.

Psychological Essentialism depends for its insights on the related philosophical doctrine of Essentialism, as developed by Putnam and Kripke, and more recently by Rey (1983, 1985). The Psychological version of the view holds that people subscribe to the existence of essences. The philosophical doctrine is that concepts actually do have essences. It is important to note that the psychological theory is independent of whether or not the philosophiocal view is correct or not. We may mistakenly believe in essences that don't exist, or similarly we may be mistaken in not believing in essences that actually do exist. It is important therefore that discussions about the nature of psychological representations of concepts (epistemology) do not become confused with discussions about the true nature of the real world (ontology) -- see Rey (1983, 1985) and Smith, Medin & Rips (1984).

How do essentialist beliefs work? According to Putnam, when a particular class of things is given a name, then by the name we mean to refer to whatever is essential to that class - whatever it may be. Hence the word "water" refers to the stuff that we call water, and anything else that has the same essential character. According to Putnam and Kripke, the first naming event "fixes" the reference of the term in a rigid way. We may discover that our current theory that water is H_2O is incorrect, and that in fact water is something else - however water will remain water regardless. Effectively, Putnam argues, we must hand over responsibility for determining the true meaning of many of the words and concepts that we use to the "experts" in society who are in charge of discovering the essential nature of the world. It follows that for many people and for many of their concepts, the concept representation must contain an empty marker for the concept definition. For example if asked what makes petrol petrol rather than anything else -- what is its essential nature -- most people will admit that they don't know, but that they believe that there is some essence, which some expert chemist could probably define. Our conceptual representation for petrol would therefore contain (along with information about its use, characteristic appearance and smell, combustability and origin) a "place-holder" for the critical essence, which would say "Ask a chemist!".

Evidence and Critique

It is undoubtedly true that our concepts may often form part of wider networks of related conceptual structures, and that in some cases these structures could be described as "theories". The similarity-based theory of concepts (as exemplified by the prototype and exemplar theories) has to rely on the statistical properties of the objects in the environment in order to account for the concepts that we develop. Given that we attend to certain dimensions and attributes of the environment, then statistical clustering on the basis of similarity would provide us with a structure of categories and subcategories with prototype-like properties (in terms of graded typicality, borderline uncertainty, lack of clear-cut definitions and lack of mutually exclusive taxonomically ordered classes). Once a child has developed such a structure, however, why should it change? Work in developmental literature (e.g. Keil and Batterman, 1984) shows that children may often start out with apparently similarity-based concepts, but that there is a trend to replace these with deeper, more theoretically relevant concepts as the child grows older. We therefore clearly need an account of the development of deeper understanding, and how conceptual structure has to change to facilitate this understanding.

Sometimes however it appears that psychologists have overestimated the depth of our conceptual understanding. Recent work by Malt (1990, 1991, 1994; Malt and Johnson, 1992) has challenged the theory-based view of artifact and natural kind concepts. Consider for example the case of water (Malt 1994). It is commonly claimed that water is a well-defined, scientifically based concept - one with a well known essence, namely H_2O. Malt collected a large number of examples of familiar liquids like coffee, beer, pond water, or swimming pool water. She separated the liquids into those that were called water in the context from which they were taken and those that were "similar to" but not actually water. Malt then asked subjects to estimate what percentage of H_2O was to be found in each one. Surprisingly, the estimates of percentage H_2O varied widely but were of no use for discriminating examples of water from other liquids. It appeared that whether something was called "water" or not, depended not only on its chemical composition, but also on much less scientific aspects, such as its current location, its source and its function for humans.

Malt's work illustrates an important point that is often overlooked in work on concepts - that the way in which we use words to describe the world is not a purely conceptual problem, but also involves the historical development of the language that is being used. There is a tension between the intuition that a word has a meaning that is derived from its semantics in a particular language, and the intuition that word meanings are direct reflections of our conceptual understanding of the world. This tension is frequently unresolved in both the minds of experimenters and in the minds of their subjects.

A recent study of my own (Hampton, 1995) further illustrates the difficulty of assuming that the "average person" has well developed theory-based concepts. In this study, subjects were given different scenario instances of a range of concepts. For example they were given different descriptions of objects and asked to say whether the object described was an UMBRELLA. The object descriptions were manipulated with the intention of independently varying a feature that should be part of the core definition of the concept class (for example that the object was designed and used for holding above one's head to keep off the rain), and a set of features that should be incidental to the category definition (for example that the object was a dome made of cloth and wire, as opposed to a flat hexagon made of plastic and wood). Core defining features were constructed on the basis of the "theory-based" model of concepts -- namely, artifacts were defined in terms of intended and actual function, and biological kinds were defined in terms of inheritance of genetic information. The expectation was that the absence of a core defining feature would bring categorization down from 100% to 0%, whereas manipulation of the incidental features would have no effect on categorization. Surprisingly, it proved extremely difficult to identify sets of features for which this pattern could be obtained. In spite of four recursive attempts to improve the materials for the experiment, weeding out poor examples, and strengthening others, it was impossible to prevent the indicental features from influencing categorization. For example, when told a story about a Zebra, the offspring of two normal zebra parents, which was given a special nutritional diet during development which led it to develop into something that looked and behaved just like a horse, two thirds of the subjects stated that it was no longer a zebra.

A re-examination of Rips' (1989) study with the metamorphosis of a bird into an insect reveals that in his data also there was an effect of the "incidental" information on the likelihood of categorization. Similarity-based effects are thus difficult to eradicate from people's categorization. It may of course be the case that subjects are poor theorists. Many people may well believe that nutritional diet can affect the "essential" nature of some organism. The results of Hampton (1995) may therefore also be taken as supporting the Psychological Essentialist theory that people do not really know what makes something fall into a certain conceptual class. A rapprochement may indeed be possible between prototype theory and psychological essentialism, if one allows that people believe in essences, but failing to have very much knowledge about them are inclined to base most of their categorization on similarity to prototype representations. The use of "essences" in people's categorization may be highly restricted in practice.

If most people have "empty slots" for representing concept essences, then what is the value of representing the essence at all? One answer was suggested by Rips (1995) who discusses a problem frequently raised in philosophy concerning the stability of concepts. If the identity of a concept is determined by all the semantic information used to represent it (as for example in the prototype view), then it becomes hard to explain how it is that conceptual differences and development occur for a single concept. This difficulty arises because if two people (or the same person at different times) have two different prototype representations, then it follows that they simply have two different concepts. It is not possible to individuate concepts in terms of the full set of semantic information they contain without coming to the conclusion that everyone has a different set of concepts, and therefore that mutual understanding is impossible. For example if you and I have different conceptual representations of the concept DAISY, then there is in principle no way of showing that they are in fact both representations of the "real" class of daisies, rather than representations of two different concepts - my concept of DAISY and your concept of DAISY. The "empty essence" slot may then be a way of adding commensurability and stability to concepts. If you and I both acknowledge that there is a real class of daisies that is defined by the daisy "essence", then we can agree that we are talking about the same class of entities, and have discussions and even disagreements about matters of fact pertaining to the class, without the problem arising of knowing whether we are talking about the same thing.

Conclusions

This chapter has reviewed four major types of model of conceptual representation. What becomes clear as one reviews the evidence for and against each model, is the variety and complexity of human concepts. There is a temptation for theorists to wish to apply their own approach to all conceptual representations. It is however most unlikely that all concepts are defined or represented in the same way. What is needed for the advance of the field is for a principled account to be given of the range of representational powers that people possess, and for a matching up of different kinds of representation with different conceptual domains. Keil (1989) has argued for a two-process view of concept acquisition and representation. On the one hand there are domain-general procedures for recording frequency and correlation information about the co-occurrence of attributes across categories. Such procedures would be good at detecting prototypes, and at storing relevant salient exemplars or subtypes of superordinate classes, and could be easily implemented in neural network learning paradigms. On the other hand, Keil argues that there are domain-specific beliefs about causal relations in different domains. Keil suggests that broad domains such as sentient beings, biological kinds and artifacts are represented with respect to quite different domain theories. These theories which must presumably have innate origins drive the child's hypothesis formation about which aspects of the world are relevant for concept formation in different domains. Without constraints on hypothesis formation, it is clear that induction of concepts from raw data is not possible. There are just too many possible ways of categorizing the world for the problem to be soluble without some strong prior constraint. Keil proposes that the child comes equipped with different constraints for different conceptual domains. If this is the case, then we should also expect differences to also emerge in adult conceptual representation across different domains. The future for the psychology of concepts may then lie in investigating such differences.

References

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

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.

Barsalou, L.W. (1987). The instability of graded structure: implications for the nature of concepts. In U.Neisser (ed.) Concepts and Conceptual Development: Ecological and Intellectual Bases of Categories (pp. 101-140). Cambridge: Cambridge University Press.

Barsalou, L.W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T. Skrull & R.S.Wyer (Eds.), Advances in social cognition, Volume III: Content and process specificity in the effects of prior experiences, (pp 61-88) Hillsdale, NJ: Lawrence Erlbaum Associates.

Barsalou, L.W. (1992). Frames, concepts, and conceptual fields. In E. Kittay & A. Lehrer (Eds.), Frames, fields, and contrasts: New essays in semantic and lexical organization. (pp. 21-74). Hillsdale, NJ: Lawrence Erlbaum Associates.

Barsalou, L.W. (1993). Structure, flexibility, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A.C.Collins, S.E.Gathercole, & M.A.Conway (Eds.), Theories of memory (pp. 29-101). London: Lawrence Erlbaum Associates.

Barsalou, L.W., & Hale, C.R. (1993). From feature lists to frames. In I.van Mechelen, J.A. Hampton, R.S. Michalski, & P. Theuns (Eds.), Categories and concepts: Theoretical views and inductive data analysis. (pp. 97-144). London: Academic Press.

Barsalou, L.W., Huttenlocher, J., & Lamberts, K. (1995). Processing individuals during categorization. (Unpublished manuscript).

Brooks, L.R. (1978). Nonanalytic concept formation and memory for instances. In E.Rosch and B.B.Lloyd (Eds.), Cognition and Categorization, Hillsdale, NJ: Lawrence Erlbaum Associates.

Brooks, L.R. (1987). Nonanalytic cognition. In U. Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual factors in categorization, (pp. 141-174). Cambridge: Cambridge University Press.

Bruner, J.S., Goodnow, J.J., & Austin, G.A. (1956). A Study of Thinking. New York: Wiley.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. Journal of Personality and Social Psychology, 35, 38-48.

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology, 12, (pp. 3-52), New York: Academic Press.

Cantor, N., Mischel, W., & Schwartz, J.C. (1982). A prototype analysis of psychological situations. Cognitive Psychology, 14, 45-77.

Cantor, N., Smith, E.E., French, R., & Mezzich, J. (1980). Psychiatric diagnosis as prototype categorization. Journal of Abnormal Psychology, 89, 181-193.

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

Chater, N., & Heyes, C. (1994). Animal concepts: Content and discontent. Mind and Language, 9, 209-246.

Chomsky, N. (1980). Rules and representations. New York: Columbia University Press.

Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-248.

Collins, A.M., & Quillian, M.R. (1972). How to make a language user. In E. Tulving and W. Donaldson (Eds.), Organization of memory. London: Academic Press.

Conrad, C. (1972). Cognitive economy in semantic memory. Journal of Experimental Psychology, 92, 149-154.

Fodor, J.A., Garrett, M.F., Walker, E.T., & Parks, C. (1980). Against definitions. Cognition, 8, 1-105.

Fodor, J.A. (1994). Concepts - a pot-boiler. Cognition, 50, 95-113.

Gelman, S.A. (1988). The development of induction within natural kind and artifact categories. Cognitive Psychology, 20, 65-95.

Gelman, S.A., & Markman, E.M. (1986). Categories and induction in young children. Cognition, 23, 183-208.

Hampson, S.E. (1982). Person memory: a semantic category model of personality traits. British Journal of Psychology, 73, 1-11.

Hampton, J.A. (1979). Polymorphous concept in semantic memory. Journal of Verbal Learning and Verbal Behavior, 18, 441-461.

Hampton, J.A. (1981). An investigation into the nature of abstract concepts. Memory and Cognition, 12, 151-164.

Hampton, J.A. (1982). A demonstration of intransitivity in natural concepts. Cognition, 12, 151-164.

Hampton, J.A. (1987). Inheritance of attributes in natural concept conjunctions. Memory and Cognition, 16, 579-591.

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. (1993). Prototype models of concept representation. In I.van Mechelen, J.A. Hampton, R.S. Michalski, & P. Theuns (Eds.), Categories and concepts: Theoretical views and inductive data analysis pp. 67-95. London: Academic Press.

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

Hampton, J.A. (1996). Conceptual combination. *In* K. Lamberts and D. Shanks (Eds.) *Knowledge, Concepts and Categories.* London: UCL Press.

Hampton, J.A., & Dubois, D. (1993). Psychological models of concepts: Introduction. In I.van Mechelen, J.A. Hampton, R.S. Michalski, & P. Theuns (Eds), Categories and concepts: Theoretical views and inductive data analysis. London: Academic Press.

Heider, E.R. (1972). Universals in color naming and memory. Journal of Experimental Psychology, 93, 10-20.

Hoffman, J. (1982). Representations of concepts and the classification of objects. In F.Klix, J.Hoffman, & E. van der Meer (Eds.), Cognitive research in psychology. Amsterdam: North-Holland.

Jardine, N., & Sibson, R. (1971). Mathematical Taxonomy. London: Wiley.

Jolicoeur, P., Gluck, M.A., & Kosslyn, S.M. (1984). Pictures and names: Making the connection. Cognitive Psychology, 16, 243-275.

Katz, J.J., & Fodor, J.A. (1963). The structure of a semantic theory. Language, 39, 170-210.

Keil, F.C. (1981). Constraints on knowledge and cognitive development. Psychological Review, 88, 197-227.

Keil, F.C. (1986). The acquisition of natural kind and artifact terms. In W. Demopoulos & A. Marras (Eds.), Language learning and concept acquisition: Foundational issues. Norwood, NJ: Ablex.

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

Keil, F.C., & Batterman, N. (1984). A Characteristic-to-Defining shift in the development of word meaning. Journal of Verbal Learning and Verbal Behavior, 23, 221-236.

Kripke, S. (1972). Naming and necessity. In D.Davidson and G.Harman (Eds.), Semantics of Natural Language. Dordrecht: Reidel.

Landau, B. (1982). Will the real grandmother please stand up? The psychological reality of dual meaning representation. Journal of Psycholinguistic Research, 11, 47-62.

Malt, B.C. (1990). Features and beliefs in the mental representation of categories. Journal of Memory and Language, 29, 289-315.

Malt, B.C. (1991). Word meaning and word use. In P.Schwanenflugel (Ed.), The psychology of word meanings. Hillsdale, NJ: Erlbaum.

Malt, B.C. (1994). Water is not H_2O. Cognitive Psychology, 27, 41-70.

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., & Ortony, A. (1989). Psychological essentialism. In S.Vosniadou & A.Ortony (Eds.), Similarity and analogical Reasoning. Cambridge: Cambridge University Press.

Medin. D.L., & Schwanenflugel, P.J. (1981). Linear separability in classification learning. Journal of Experimental Psychology: Human Learning and Memory, 7, 355-368.

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

Medin, D.L., & Shoben, E.J. (1988). Context and structure in conceptual combination, Cognitive Psychology, 20, 158-190.

Miller, G.A., & Johnson-Laird, P.N. (1976). Language and Perception. Cambridge Mass: Harvard University Press.

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.

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

Osherson, D., & Smith, E.E. (1982). Gradedness and conceptual conjunction. Cognition, 12, 299-318.

Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 77, 353-363.

Putnam, H. (1975). The meaning of 'meaning'. In Mind, language and reality, volume 2: Philosophical papers. Cambridge: Cambridge University Press.

Randall, R.A. (1976). How tall is a taxonomic tree? Some evidence for dwarfism. American Ethnologist, 3, 543-553.

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

Rey, G. (1985). Concepts and conceptions: A reply to Smith, Medin & Rips. Cognition, 19, 297-303.

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. (1995). The current status of research on concept combination. Mind and Language, (in press).

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.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 8, 382-439.

Ryle, G. (1951). Thinking and language. Proceedings of the Aristotelian Society (Supplementary Series), 25, 65-82.

Shanks, D.R. (1996). Representation of categories and concepts in memory. (this volume).

Shepard, R.N., Hovland, C.I., & Jenkins, H.M. (1961). Learning and memorization of classifications. Psychological Monographs, 75, 1-42.

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

Smith, E.E., Medin, D.L., & Rips, L.J. (1984). A psychological approach to concepts: Comments on Rey's "Concepts and Stereotypes". Cognition, 17, 265-274.

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. (1994). Rule-based and similarity-based categorization. Memory and Cognition, 22, 377-386.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), Organization of memory. London: Academic Press.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. Psychological Review, 89, 123-154.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. Journal of Experimental Psychology: General. 113, 169-193.

Wittgenstein, L. (1958). Philosophical Investigations. New York: Macmillan.

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