Categories, Prototypes and Exemplars

James A. Hampton, City University London

Chapter 8 for the Routledge Handbook of Semantics, Ed. Nick Reimer (forthcoming)

INTRODUCTION

This chapter describes an approach to theorizing about the meaning of words that is primarily based in the empirical research methods of cognitive psychology. The modern research tradition in this area began with the notion introduced by Tulving (1972) of *semantic memory*. Tulving pointed to an important distinction between the memories that each individual has of their own past (which he termed *episodic memory* – memory for events and episodes of experience), and the general conceptual knowledge of the world that we all share, which he termed *semantic memory.* There is some ambiguity about just how broadly the notion of semantic memory should be taken. For example does it include all facts that you know which are not based on actual experiences, or should it be restricted to conceptual knowledge about what *kinds* of things there are in the world and their properties? Nonetheless the central contents of semantic memory are quite clear. The semantic memory store contains the concepts that enable us to understand and reason about the world, and as such it provides the knowledge base that underpins the meanings of utterances and individual words in any language. Knowing that a bird is a creature, or that chemistry is a science involves a conceptual knowledge network where cultural and linguistic meanings are represented: semantic memory is a combination of mental dictionary in which words are given definitions and a mental encyclopaedia in which general information concerning the referent of the word is stored.

There is general agreement that semantic memory is largely separate from episodic or other forms of memory (such as memory for motor actions). In particular, people may suffer severe forms of amnesia while retaining their production and comprehension of language.

Semantic memory models of the 1970s were based on two main theoretical ideas. One was to consider semantic memory as a form of network. Collins and Quillian (1969) developed a structural model of semantic memory in which concepts were nodes in a network, joined by labelled links. For example the word BIRD would be linked by a *superordination* "Is a" link to ANIMAL, and by a *possession* "Has a" link to FEATHERS. In the same way the word would be linked to a range of the properties that it possessed, classes to which it belonged, and subclasses that it could be divided into. Figure 1 shows an example for BIRD.

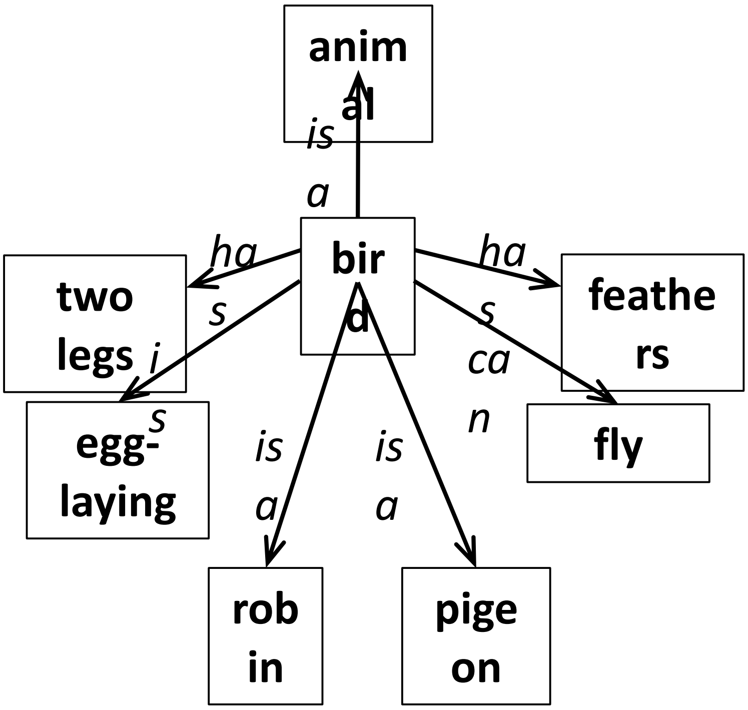

Figure 1. A semantic network representation of the meaning of BIRD.

The network provides a neat representation of how the meaning ofkinds and their properties could be inter-related. Birds are partly characterised by their possession of feathers, while feathers are partly characterised by being a property of birds. The model also has the advantage of providing an economical way of storing a large amount of information. A property only needs to be stored in the memory structure at the most general level at which it is usually true. It is not necessary to store the fact that birds have skin or that they reproduce, since these properties are true of animals in general. If questioned about whether birds reproduce, the memory system would retrieve the relevant links connecting bird to reproduction in the following way:

1) A Bird is an animal
2) An animal is a biological organism
3) Biological organisms reproduce

from which the inference would be drawn that birds reproduce.

Evidence for the semantic network model was in fact weak and the model was quickly superseded, although it has had an extraordinarily enduring presence in the text books. The primary evidence was based on reaction time measures for either judging category membership (A robin is a bird) or property possession (A robin has feathers). Collins and Quillian (1969) reported that response times to true sentences increased with the number of inferential links that needed to be retrieved. Thus “A robin is a robin” was faster than “A robin is a bird” which was faster than “A robin is an animal”. The model made predictions based on a search and retrieval process, whereby information was already present in the network, and the time required to retrieve it could be used as an index of the network’s internal structure.

Difficulties with the model arose immediately with the response times for false sentences. In the case of false statements such as “A robin is a fish”, the more intervening links there are in the network, the *faster* is a correct rejection since people are delayed by similarity between subject and predicate terms for false sentences. For example “A robin is a fish” is slower to falsify than “A robin is a vehicle”. It was therefore necessary to introduce a more complex search algorithm, involving the idea of spreading activation. Collins & Loftus (1975) proposed that activation would start at the subject and predicate nodes and then spread with diminishing strength through the network. Dissimilar nodes would lead to a rapid “false” decision as the two streams of activation died away without entering the same link. However activation of similar nodes, being in the same part of the network, would lead to retrieval of a common path linking them. This path would then have to be checked to see if it warranted the inference that the sentence was true. Slow false responses for close items were explained by the need to do this checking.

Further evidence against the network idea was not long in appearing. Smith, Shoben and Rips (1974) demonstrated how 3-level hierarchies of terms could be found in which the distance effect was reversed. For example “A chicken is a bird” was slower to verify than “A chicken is an animal”. Since the model requires that the latter is based on an inference from chickens being birds and birds being animals, there was no way to explain this type of result.

An additional problem, familiar in semantic theories, was the question of whether nodes corresponded to words, or to word senses. Should the property node “has four legs” be attached to both horses and chairs, or does the difference in the kind of leg involved require different nodes to reflect each sense of ‘leg’? Network theorists did not attempt to address these questions, leaving the model very limited in its scope.

Smith et al. (1974) proposed an alternative model based on the notion of semantic features. Semantic features as originally conceived were aspects of a word’s meaning. Features like gender and animacy play a role in explaining syntactic acceptability, while linguists also analysed semantic fields such as kinship terms in terms of their featural components.

Within structuralist linguistics features are typically defined as having three possible values. For example Number could be singular, not singular (i.e. plural) or unspecified . In Smith et al.’s model however the notion was greatly broadened to include more or less any property as a semantic feature. Having a red breast would be a feature of robins, and having a trunk a feature of elephants. Feature models aim to capture the meaning of a word in a very similar way to the network model, but instead of providing a network of connected links, each word is simply represented by a list of its features. Thus BIRD would be represented as in Figure 2.

BIRD

- ○ **is alive**
- ○ **flies**
- ○ **has feathers**
- ○ **has a beak or bill**
- ○ **has wings**
- ○ **has legs and feet**
- ○ **lays eggs**
- ○ **has just two legs**
- ○ **builds nests**
- ○ **sings or cheeps**
- ○ **has claws**
- ○ **is very lightweight**

Figure 2. A feature representation of the meaning of BIRD.

Rather than using search and retrieval as the process underlying semantic verification, the feature model made a different assumption. Focussing on categorization decisions (e.g. a robin is a bird), the model proposed that a decision is based on a comparison of the set of features that defines ROBIN and the set of features that defines BIRD. If robins possess all of the features that are necessary for being a bird, then the sentence will be judged true. Otherwise it will be false.

To recap, when people judge category statements, they are faster to say True when the two words are more similar, and faster to say False when the two words are more dissimilar. To capture this result, Smith et al.'s model proposed two stages in a decision process. In the first stage, overall featural similarity was computed in a quick heuristic fashion. If the result showed either very high similarity or very low similarity, then a quick True or False decision could be given. However if the result based on overall similarity was inconclusive a second stage was required. An inconclusive first stage result would mean that a true category member lacked some of the characteristic features of the category (e.g. flightless birds such as OSTRICH or PENGUIN), or that a non-member possessed some of those features (e.g. flying mammals such as BAT). To deal with these slow decisions, the second stage required the individual defining features of the category (e.g. BIRD) to be identified and checked off against the features of the possible category member.

Smith et al.'s model was no more successful than Collins and Quillian's, although like the former model it has had great longevity in the literature. The first difficulty with the model was that Rosch and Mervis's work on prototype concepts (Rosch & Mervis, 1975) revealed that many words did not have a clear set of defining features that could be appealed to for a second stage decision. The second was that the response time results could be explained more parsimoniously in terms of a single similarity computation.

For example, Hampton (1979) showed that when people make erroneous categorization decisions, they tend to be very slow. According to Smith et al.'s model, errors should only arise in the rapid heuristic first stage – for example when someone agrees that a bat is a bird on the basis of their similarity without checking the defining features in stage 2. The slow second stage should be more error free, given that the defining features are carefully checked off. In fact, most semantic categories have borderline regions where response times become very slow, and people's responding becomes less consistent (McCloskey & Glucksberg, 1978). This pattern of data is more consistent with the Prototype model described below.

The feature model also had little explanation to offer for the verification of properties. The proposal that categorization depends on feature checking would imply that properties should be verified faster than category membership judgments. In fact, the reverse is the case. Hampton (1984) measured verification times for two kinds of sentences: category statements such as Robins are birds and property statements such as Robins have wings. Sentences were equated for their production frequency in a feature listing task. The category statements were consistently faster to verify, making it implausible that people judge that a robin is a bird by first verifying that it has wings (and other defining features).

This brief historical overview of semantic memory research serves to set the background for more recent theories of how people represent word meaning. Semantic networks have proved the inspiration for very large scale network analysis of meaning using statistical associations (see e.g. Chapter 7). Feature-based models have been the inspiration for schema-based representations of concepts and meanings, including recent versions of prototype and exemplar models.

WORD MEANING AND SEMANTIC CATEGORIES

A function of word meaning, much studied in psychology, is to categorize the world into labelled classes. Although semantics is concerned with the meaning of all language, research in psychology has focused in a rather narrow fashion on nouns, and in particular on semantic categories. The reason for this is that categories such as Sport, Fruit or Science provide a rich test bed for developing theories of concepts and meaning. Notably, (a) they are culturally specific to a degree but also tied to objective reality, (b) they have familiar category "members", subclasses like Tennis, Lemon or Physics, which can be used in experiments on categorization and reasoning, and (c) people can introspect on the basis they use to classify, and can describe general properties of the classes.

With more abstract concept terms like RULE or BELIEF, it is hard for people to reflect on the meaning. Hampton (1981) found that people had difficulty in performing each of the tasks that would allow the construction of a prototype model for abstract concepts. By comparison many studies have shown that people can readily generate both exemplars (i.e. category members) and attributes (i.e. properties) for categories such as Sport or Fruit. These domains therefore provide a good way to test just how the meaning of the category concept is represented psychologically.

Prototypes

The idea of representing a conceptual meaning with a prototype owes much to the pioneering work of Rosch and Mervis (1975). A prototype represents a kind in terms of its most common and typical properties. However no individual property need be true of the whole kind (although some may be), so that belonging to the category simply involves possession of a sufficient number of such properties. Exemplars will also differ in typicality as a function of the number of such properties they possess. More broadly a prototype concept is one whose reference is the set of all exemplars whose similarity to a prototype representation is greater than some threshold criterion.

In a series of highly influential papers Rosch and Mervis presented a large array of empirical work showing that many concepts underlying semantic categories had a prototype structure. A standard methodology to show this structure has evolved, and typically uses all or some of the following steps (each with a different group of participants):

A. People generate exemplars of the category. For example for SPORTS, participants list all the sports that they can think of.
B. Participants judge the typicality of each of the exemplars. By typicality is meant the degree to which the exemplar is representative of or typifies the category as a whole. For example football and tennis are considered typical as Sports.
C. The exemplars are listed together with other items from the same domain (e.g. other human recreational activities) and participants make category membership judgments about each item.
D. Participants generate attributes for the category. They list what sports tend to have in common, that differentiates them from other types of thing.
E. Participants judge the “importance” or “definingness” of each of the attributes for the meaning of the category. For example how important is the attribute “is competitive” for the category of sport?
F. All the exemplars and attributes generated with a certain minimum frequency are placed as rows and columns of a matrix, and participants complete the table with judgments of whether each exemplar possesses each attribute.

When all this has been done, the analysis can then proceed. Prototype structure is revealed in four ways.

1) The category boundary is found to be vague in (C). There are a significant number of activities which are borderline cases where people cannot agree about the categorization.
2) There are systematic variations of typicality across category exemplars in (B), which correlate with frequency of generation in (A) and other cognitive measures.
3) Just as there is no clear set of category members, so there is no clear set of category attributes in (D) and (E). Attributes can be ranged along a continuum of definingness, and people will disagree about whether some attributes should be counted or not as part of the concepts meaning.
4) Most importantly, when the matrix of exemplar/attribute possession from (F) is examined, no definition can be offered such that all of the category members and only the category members possess some fixed set of attributes. Being a sport is not a matter of possessing a set of singly necessary and jointly sufficient defining attributes. Rather there is a clear correlation between the number (and definingness) of the attributes that an exemplar

> possesses and the degree to which it is typical of the category and/or the degree to which people agree that it belongs.

Using this type of methodology in the years following Rosch and Mervis (1975), prototype models were found across a range of different domains, including speech acts like lying, psychiatric categories, and personality ascriptions, as well as artifacts, human activities and folk biological kind categories.

There has been much debate about the validity of the prototype model as a theory of concepts. It is therefore worth clarifying the notion. First, the theory should not be confused with the operational methodology used to provide evidence of it. It is not supposed for example that the mind contains a list of attributes in the verbal form in which they are generated. Clearly meaning has to be grounded (see Chapter 9) in experiential sub-symbolic levels of cognition, so it is unhelpful for a psychological model to give the meaning of one word simply in terms of others unless there is a primitive base of terms that are defined non-verbally. Second, while it is true that prototype theory defines the extension of a term in terms of similarity (number and weight of matching versus mismatching attributes), it is not wedded to any particular theory of similarity per se. One proposal is that degree of category membership and typicality can be associated with distance from the prototype in a similarity space. However there are important ways in which similarity does not map into proximity in a space which undermine this proposal. A space assumes that the same dimensions are relevant for all similarity comparisons, but this is clearly not the case – A and B may be similar for different reasons than B and C. In addition, the prototype that represents the category has to be represented at a higher level of abstraction than the prototypes that represent its members. It is for this reason that the prototype is *not* to be equated with the most typical exemplar. The concept of Bird is not equivalent to the concept of Robin. There are attributes of robins (for example its coloured breast) which do not figure in the more abstract concept of Bird, so that we can agree with "A robin is a bird" and disagree with "A bird is a robin".

The key proposal of prototype theory is that meaning is represented in the mind through an idealised general representation of the common attributes of the extension (the referents) of the term. It is this information that people are able to access when they generate lists of attributes or judge how central an attribute is to the meaning of a term. The reference of a term is then determined by similarity to this prototype representation. This mechanism for determining reference provides prototype theory with an advantage over many other accounts in that it directly explains the vagueness and imprecision of meaning. The vagueness of language has been the source of much debate in semantics (Keefe & Smith, 1997), as it presents a serious challenge to the determination of truth for propositions and combinations of propositions in natural language. Traditionally, vagueness has been identified with the problem of determining the truth of statements using scalar adjectives such as TALL or BALD, where on the one hand it is clear that there is some height at which "X is tall" turns from False to True, but on the other hand it seems impossible to identify any particular height as being the critical value except through arbitrary stipulation. Prototype theory shows that similar problems of vagueness can arise with the multi-faceted concepts that underlie noun terms.

Vagueness in noun categories can be expressed in prototype theory by proposing *degrees of membership*. A tomato is a fruit to a certain degree. The relation between graded membership and typicality has also been a source of confusion. According to the theory, as an item becomes increasingly dissimilar from a category prototype, first its typicality declines, although it remains a clear member of the category. Then as it reaches the boundary region of the class, both typicality and degree of membership will decline, until it is too dissimilar to belong to the category at all. It is probably confusing to talk of the typicality of items that are NOT members of a category, although in fact Rosch and Mervis did ask their participants to rate typicality for such items, and people did not apparently object.

Evidence for prototypes

The primary evidence for prototypes comes from studies using the procedure outlined above. Where a meaning carries many inferences (attributes such as that if X is a bird, X can fly), many of which are probabilistic (not all birds actually fly), and where membership of the category is vague at the boundary, then a prototype is likely to be involved. Paradoxically, however, these are not necessary features of a prototype meaning. If there is a cluster of attributes which are strongly correlated, and if the world happens to contain no cases that would lie near the boundary in terms of similarity, then it is possible that a prototype representation would in fact be compatible with a conjunctive definition and no borderline vagueness. Consider the example that has been used so far of Birds. Birds are the only feathered bipedal creatures, and since their evolutionary ancestors among the dinosaurs became extinct long ago, there are no borderline cases within the folk taxonomy of kinds (which is our primary concern as psychologists). So BIRD has a clear-cut definition – 'feathered bipedal creature' – and no borderline cases. But there is no reason to suppose that people represent them differently from bugs, fish and reptiles, which are much less clearly represented as concepts.

A classic study by Malt and Johnson (1992) demonstrated the prototype nature of artifact concepts. They constructed descriptions of unfamiliar objects that might either have the appearance but not the function of a given artifact like a BOAT, or alternatively the function but not the appearance. They were able to show that having the correct function was neither sufficient nor necessary for something to be judged to belong in the category. The absence of a set of singly necessary and jointy sufficient defining attributes is a hallmark of prototype representation.

An immediate worry about the method for finding prototypes is that the outcome may result from averaging and summing across individuals who themselves may have clearer definitions of their meanings. If the linguistic community contained three different clearly defined ideas of what Sport means, then the result of combining the ideas in the procedure will look like a prototype. McCloskey and Glucksberg (1978) were able to show that this is not the case. They asked a group of participants to make category membership judgement for a range of categories. In the list of items were many borderline cases, and this was shown in the degree to which people disagreed about them. The participants were then asked to return some weeks later and repeat their judgments. If it was the case that each individual had their own clearly defined category, then the judgments should have shown high consistency between the first and second occasion. In the event however, there was a high level of inconsistency for those same items about which people disagreed. There were some stable inter-individual differences, but the main result was that vagueness exists within the individual rather than just between individuals.

Research on prototypes has also demonstrated that people's use of language can often be shown to deviate from logical norms in ways that are readily explained by the theory. The theory uses an internalist account of semantics (see Chapter 2), whereby meaning is fixed by the mental contents of the representation of a concept in memory. In addition, by depending on similarity to determine the reference of a term, the door is opened to various forms of “reasoning fallacy”. Similarity is a relatively unconstrained measure, since the basis on which similarity is calculated can shift depending on what is being compared. North Korea may be similar to Cuba in terms of politics, while Cuba is similar to Barbados in terms of climate and location, but there is little or no similarity between North Korea and Barbados. This shifting of the basis for similarity can lead to intransitivity in categorization as Hampton (1982) showed. If categorization has the logical structure of class inclusion, then it should be transitive. If A is a type of B, and B a type of C, A should be a type of C. In the study people were happy to agree that chairs were a typical type of furniture. They also agreed that ski-lifts and car-seats were kinds of chair, but they did not want to call them furniture. While in logical taxonomies the “Is a” relation is transitive, in natural conceptual hierarchies it is not. Similarity is the culprit. In deciding that a chair is a kind of furniture, people are focused on how well chairs match the attributes typical of furniture. Then when deciding if a car-seat is a kind of chair, they focus on the attributes typical of chairs. As the basis of determining similarity shifts, so the intransitivity becomes possible.

As illustration of the power of the prototype theory to account for a wider range of phenomena, consider the following two reasoning fallacies. Tversky and Kahneman (1983) introduced the famous conjunction fallacy. They described the case of Linda who was involved in liberal politics in college. Participants were then given various statements to rank in terms of their probability. They consistently considered “Linda is a feminist bank teller” as more likely than the plain “Linda is a bank teller” even though the first is a subclass of the second. The results were explained in terms of *representativeness*. People judge probability on the basis of similarity – in this case on the basis of the similarity between what was known about Linda and the two possible categories she was compared with. The subclass relation between bank tellers and feminist bank tellers was never considered.

The second fallacy was reported by Jönsson and Hampton (2006) as a phenomenon which we called the *inverse* conjunction fallacy. As with the conjunction fallacy, the issue concerns fallacious reasoning about subclasses. In our study we gave people two universally quantified sentences such as:

All sofas have backrests

All uncomfortable handmade sofas have backrests

Different task procedures were employed across a series of experiments, but the general result was that people considered the first sentence more likely to be true, even though the second was entailed by the first. Hampton (2012) argues that people are thinking “intuitively” in terms of intensions. Backrests are a typical property of sofas, but less typical of uncomfortable handmade sofas. In spite of the universal quantifier, this difference in property typicality drives the judgment of likely truth. (Similar effects also occur in inductive reasoning where dissimilarity can undermine people’s assessments of the truth of apparently certain inferences.)

Standard semantic theories have difficulty with accounting for these results. Prototype theory explains (and in fact predicts) their occurrence. The effects of conjunction on prototype representations (as in feminist bank teller, or handmade sofa) have been widely studied (see Hampton, 2011). The meanings of the two terms interact at the level of individual attribute values so that the meaning of the conjunction is no longer determined in a simple compositional way (see Chapter 26). For example, in the case of the sofa, a backrest implies comfort, whereas the modifier "uncomfortable" implies the opposite. The interaction between these conflicting features throws doubt on whether the backrest will still be there. **Context effects and prototypes**

One possible source of the variability seen in people's prototypes may come from context (see Chapter 12). Clearly, the notion of "sport" is likely to shift in the context of a kindergarten sports day, the 2012 London Olympics, or a country house weekend in Scotland. Intuition suggests that there is some common core to the different contextually shifted meanings, but prototype theory would suggest that rather than still retaining some common definitional core, each meaning in fact involves a shift away in a different direction from a common prototype, to the point where there is very little in common across the different senses of "sport".

Studies have shown a strong influence of various contextual factors on how people judge typicality of instances. For example Roth and Shoben (1983) manipulated the scenario in which a concept appeared (e.g. "the bird crossed the farmyard" or "the trucker drank the beverage"). Instances typical to the context (chickens or beer) were boosted in processing speed. In another study Barsalou and Sewell (1984) showed that if people were asked to adopt the point of view of another social group (e.g. housewives vs farmers), then their typicality judgments would completely shift. Remarkably, students' agreement about typicalities from the adopted point of view was not much lower than their agreement about their own (student's) point of view.

Another study notably failed to find any effect of context on categorization. Hampton et al. (2006) provided participants with lists to categorize containing borderline cases (such as whether an avocado is a fruit or whether psychology is a science). Participants were divided into different groups and given different contextual instructions. For example in the Technical Condition, they were asked to imagine they were advising a government agency such as a Sports Funding Council on what should be considered sport. In the Pragmatic Condition in contrast they were asked to imagine that they were devising a library index that would place topics of interest in categories where people would expect to find them. A Control Condition just categorized the items without any explicit context. Various measures were taken of categorization within each group, including the amount of disagreement, individual consistency across a period of a few weeks, the size of the categories created, the correlation of category probability with an independent measure of typicality, and the use of absolute as opposed to graded response options. Across six categories and two experiments there was very little evidence that the instructions changed the way in which people understood or used the category terms. Overwhelmingly the likelihood that an item would be positively categorized was predicted by the item's typicality in the category in an unspecified context with near perfect accuracy.

It is debatable whether there are in fact stable representations of concepts (and hence word meanings) in memory, or whether concepts are only ever constructed within a particular context. It is fair to say that every use of a concept in language will involve a contextual component – there is no direct window into what is represented. But at the same time it is reasonable to hypothesize that there is some permanent information structure in memory on which the context operates. Prototype representations have the flexibility to allow for contextual modification. On the other hand, exemplar models (see below) provide even greater flexibility with each context driving the concept representation through retrieval of a set of similar previously encountered contexts.

Critiques of the Prototype Theory of meaning

Critics of prototype theories of concepts were not long in coming to the fore. Rey (1983) pointed out that all of the problems that have been identified with descriptivist or internalist semantics as an account of concepts apply equally to the prototype theory. If conceptual (or meaning) content is equated with the information represented in a person's mind, then it becomes difficult to provide an account of truth. It does not seem right to say that a person who believes that snakes are slimy (which in fact they are not) is speaking the truth when they utter the statement "snakes are slimy", even though the meaning of snake for them includes its sliminess. If the truth/falsehood of these kinds of sentences were entirely determined analytically in terms of meanings, this would lead to an alarming solipsism. Another issue is the unlikelihood that two people will ever share the same meaning for a word, given the instability and individual variation seen in prototypes, so it is then hard to explain successful communication or disagreement about facts. Once again, each individual is in their own solipsistic world of meaning.

Handling these critiques leads into complicated areas. Perhaps the best way through the maze is to point out that in possessing a meaning of a word, the language user is not the person in charge of what the word means. Their prototype has to be connected to two sources of external validation. First, the physical and social world place constraints on conceptual contents. The person with the false belief about slimy snakes will change that belief when they first touch one. Second, people's meanings are constrained by the group of language users to which they belong. There are normative rules about the use of words which get enforced to greater or lesser extent in the course of everyday conversation and language exchange. More importantly, a person should be willing to accept the "defeasibility" of their conceptual meanings. They should be happy to defer to reality or to social norms when shown they are out of line.

Within psychology, there have been two further developments from the first prototype model proposed by Rosch and Mervis (1975). Ironically perhaps, they have taken the field in two opposite directions, either increasing the representational power of the model, or reducing it. The argument In favour of increased representational power was first made in an influential paper by Murphy and Medin (1985). They argued that lexical concepts are not simply classification devices based on similarity, as the prototype account seemed to suggest. Rather, concepts provide a means of understanding and predicting the world that can incorporate deeper theoretical structures. Rather than classifying an instance in the category to which it bears maximum similarity, they suggested that people classify instances in the category that best explains its observable properties. Development of this idea suggests that prototypes are in fact knowledge schemas, inter-related networks of attributes with causal explanatory links between them. For example, birds' light-weight bones, together with their wings ENABLE them to fly, which CAUSES them to perch on trees, and ENABLES them to escape predators. This "theory" notion provides an account of how we reason with concepts. There are numerous demonstrations of the power of this approach – particularly in the developmental literature where it has been shown that children quickly learn to go beyond perceptual similarity in forming conceptual classes. In effect, words have to serve many purposes. One is to provide simple names for the things in the world around us so that we can easily communicate about them. Another is to provide the means of cultural transmission of complex ideas that have taken centuries to refine. Words like “mud” or “bug” are of the former kind, and are best modelled as prototypes. Words like “nitrogen” or “polio” refer in a different way, through their role in a deeper theory of the world and its nature.

The alternative development from prototypes has been to reduce the representational assumptions and propose that meanings are represented by sets of stored individual occurrences or exemplars. For example the meaning of BIRD would be represented through storing memories of individual instances of actual birds such as robins, sparrows, and penguins. One clear reason behind this approach is that it captures the way in which children learn language. Only rarely is a word learned by reference to a definition given by an adult or other source. Most of our words are just acquired through hearing them spoken, or reading them in text, and using the context of their use to arrive at the meaning of the word. As more and more contexts are observed, so the meaning becomes more clearly defined. However it is not necessary to assume any analysis of the meaning into attributes or schema representations. Storing individual occurrences in memory is sufficient to explain the development of an appropriate understanding of the meaning of the word.

Exemplar theory for word meanings

Exemplar models in psychology began with Medin and Shaffer (1978) from which Nosofsky (1988) later developed the best known model – the Generalised Context Model or GCM. These models were developed to explain the results of experiments in which participants were taught novel classifications of more or less artificially constructed perceptual stimuli. The advantage of such lab-based experiments is that the experimenter has full control over exactly how the stimulus classes are defined, and what learning is provided. Typically, a participant is exposed to a number of repetitions of a learning set, classifying instances in the set as in class A or class B, and receiving corrective feedback on each trial. They are then tested without feedback on a transfer set including new instances that have not previously been seen. Models are tested for their ability to correctly predict the probability that participants will endorse these new instances as an A (or a B).

The relevance of such models for lexical semantics is that they represent a laboratory model of one way in which people may learn new concepts. Hearing a term used on a number of occasions, the speaker then generalises its use to new occasions. Of course there are many differences between the laboratory task and learning in the wild. Word learning is often achieved without explicit feedback, and most lexical categories are not set up in a way that divides up a given domain into contrasting sets. However, in response, variants of exemplar models have been devised that incorporate unsupervised learning (i.e. learning without error correction) and probabilistically defined classes.

The first advantage of exemplar models over prototype representations is that there is no information loss. If every exemplar and its full context is stored in memory, then not only the central tendency of a class (e.g. its idealised member) can be computed, but also the variability within the class. (Variability can only be captured within prototype theory by the fixing of the level of similarity that is required for categorization. Highly variable classes, such as furniture would have low similarity criteria, while low variability classes such as butterflies would require a high threshold for similarity.) Because all exemplars are stored it is also possible with an exemplar representation to retrieve information about the co-occurrence of individual attributes. Small birds tend to sing, while large birds tend to make raucous calls. Prototype representations do not capture these correlations within the category, since size and type of call are each represented as independent attributes.

A second advantage is that it is possible to represent classes that are not distributed around a single central point. Prototype models assume that there is a central prototypical representation, and that all instances are classified according to their similarity to this representation. In a semantic space, this means that the model assumes that the category boundary is spherical (or hyper-spherical in more than three dimensions). But exemplar models allow that a conceptual category may have more than one similarity cluster within it. For example, sexually dimorphic creatures like pheasants form two similarity clusters around the typical male and the typical female. A creature that was an "average" of these two forms would not be a typical pheasant, even though they would be at the centre of the class. Another example from the literature is the case of spoons. Small metal spoons and large wooden spoons are each more typical than small wooden or large metal spoons. Exemplar models have no difficulty with this, since each encountered spoon is represented and the correlation of size and material is retrievable from the stored cases.

A third advantage of exemplar representations is that they incorporate frequency effects. The more common a given exemplar, then the stronger will be its influence on the categorization of others. Prototype abstraction will also be sensitive to frequency – for example the most frequent size or most frequent color will tend to be the most typical. However the frequency of given combinations of features is lost in the process of prototype abstraction.

Much of the research on exemplar models is of little relevance to lexical semantics. There are however some interesting results in support of an exemplar approach to word meaning. Gert Storms and colleagues in the Concat group at the Katholieke Universiteit Leuven have compiled a large database of semantic categories (De Deyne et al., 2008). The database includes a range of biological, artifact and activity categories. The norms provide (among other things) data on how frequently a word is generated as an exemplar of a category, how typical, imageable and familiar it is, what attributes are considered as relevant to category membership with their frequency of generation and rated definingness, and which exemplars possess which attributes. There are also pair-wise similarity judgments for exemplars across all categories.

Using these data, the Leuven group have been able to test prototype and exemplar ideas with data that are much closer to the concerns of lexical semantics. A recent review by Storms (2004) provides a useful summary. Storms first explains that in contrast to the presentation of perceptual stimuli in a laboratory experiment it is not clear just how to interpret the notion of an exemplar in semantic memory. As described above, the key issue concerns whether categorization and typicality within categories is determined by similarity to the prototype attribute set (the so-called *independent cue model*) or whether it is determined by specific similarity to other exemplars within the category, in which case relational information is also involved (the *relational coding* model). A good fit to a category is not just a question of having enough of the right attributes, it also involves having the right combinations of pairs, triples etc. of attributes. It is the involvement of this relational information that provides the sub-clustering within a category that is characteristic of exemplar representations.

Storms (2004) lists three sources of empirical evidence for exemplar models of lexical semantic concepts. In each case the question relates to whether category typicality (and categorization probability) declines in a smooth continuous fashion with distance from a central prototype, or whether there is evidence for deviations from this pattern. (Full references can be found in Storms, 2004)

A first set of tests relates to the question of Linear Discriminability. According to the prototype model, concepts in semantic memory should be linearly discriminable (LS) from each other on the basis of a simple additive combination of the available attributes. In effect category membership of an item is based on seeing whether the item has a sufficient number of the relevant category attributes, and only categories that have this structure can be represented with prototypes. Exemplar models are less constrained since weight can be given to certain configurations of features, as in the case of the metal and wooden spoons above. To test the models, researchers taught people artificial categories that either respected the LS constraint required by prototype models, or were non-linearly discriminable. Initial research found that when overall similarity was held constant, categories that respected the LS constraint were no easier to learn that those that did not, although evidence for an advantage of LS categories has since also been reported.

These studies all used artificial category learning methods. Ruts, Storms and Hampton (2004) were the first to investigate the issue using data from semantic categories. The Concat norms were used to map category exemplars into a multi-dimensional semantic space. Proximity in the space represents similarity between exemplars, so that semantic categories form clusters like galaxies in the space. Four different spaces were constructed for each domain of categories using from 2 to 5 dimensions to capture the similarity structure with increasing accuracy. Prototype theory was then tested by seeing whether categories could be distinguished from each other within the space by defining a plane boundary between them, as required by the LS constraint. In the case of biological kinds like insects, fish and birds, the categories were easily discriminated even in the low dimensionality spaces. However pairs of artifact kinds like cleaning utensils versus gardening utensils, or clothing versus accessories were not discriminable, even in the highest dimensional space. Ruts et al. concluded that prototypes were sufficient for representing biological kinds, but that artifact kinds did not respect the LS constraint, and so exemplar representations were more appropriate for those kinds.

In a second set of studies demonstrating exemplar effects in semantic memory, Storms and colleagues used the same attribute by exemplar matrices to explore whether the prototype or exemplar model provided a better prediction for four different measures of category structure. Four predictions based on the prototype model were created by using different ways of weighting the attributes in the matrix to create a sum of weighted attributes possessed by each exemplar. Predictions from the exemplar model were created by first rank ordering the exemplars in the category in order of frequency of generation to the category name. (For example "apple" might be the highest frequency exemplar generated to the category name "fruit"). Twenty-five different predictors were then created by measuring average similarity of each exemplar to either the highest one, the highest two or up to the highest 25 exemplars in the list. Finally the different predictors were correlated with four measures of category structure: rated typicality, categorization time, exemplar generation frequency and category label generation frequency. Overall the results clearly favored the exemplar model. The optimum number of exemplars involved was between 7 and 10. Individual average similarity to the top 10 exemplars in a category was generally a better predictor of the different measures than was similarity to an abstracted prototype.

The final source of evidence for exemplar effects is also from the Storms group in Leuven. Two studies looked at how people categorize unknown fresh food products as either fruits or vegetables. Thirty exotic plant foods were presented on plates to 20 participants, and were categorized as fruits or vegetables. Storms and colleagues compared prototype and exemplar models predictions of the resulting categorization probabilities. Prototype predictions were based on ratings of the degree to which each instance possessed the most important features of each category. Exemplar predictions were based on ratings of similarity of each instance to the eight most frequent exemplars of fruits and vegetables respectively. The results showed that the two models did more or less equally well, but that, interestingly, a regression model using both predictors taken together gave a significantly improved fit. In other words both possession of the right attributes *and* similarity to the most common exemplars contributed to the likelihood of categorization. For the full story of how people represent novel fruits and vegetables, see the discussion in Storms (2004).

Critique of exemplar models

There have been a number of issues raised with exemplar models, but I will briefly focus on those that are most relevant to semantic memory. The first concerns what is taken to be an *exemplar*. In the classification learning literature there is evidence that each presentation of an individual exemplar is stored, so that there is no account taken of the possibility that it might be the same individual seen on each occasion. For lexical semantics, it is more likely that one should interpret the evidence for exemplar effects in terms of what Heit and Barsalou (1996) called the *instantiation principle*. In a hierarchy of lexical concepts, such as animal, bird, eagle, the principle suggests that a particular level such as bird is represented as a collection of the concepts at the next level below. So birds are represented by eagles, sparrows and robins, while robins are represented by male robins and female robins. Below this bottom level (i.e. where there are no further familiar sub-divisions of the taxonomy) it is unclear whether or not individual exemplars (meaning actual experiences of an individual in a particular situation) are influential.

A related criticism is the problem of how people learn about lexical concepts that they never experience at first hand. How do we learn to represent the meaning of words like electron, nebula or aardvark? There must be a route to learning meanings that does not involve direct experience of individuals, since, though I have a (rough) idea of what an aardvark is, I don't recall having ever met one. A combination of pictures and written and spoken communication has provided me with all that I know about the concept.

Future Directions

Psychological studies of lexical semantics form a bridge between different research traditions, and so are well placed to attempt an integration of different sources of evidence. For psychology, the development of mathematically well-defined models of category learning has tended to sacrifice ecological validity for laboratory precision. It is important in the future that the models turn their attention to the different ways in which people acquire categorical concepts in real life. Much of learning, whether in school, college or in apprenticeships involves the development of concepts – ways of classifying experience, events or objects which provide one with predictive understanding of the world. Certainly there are cases where concepts are learned through experience with exemplars – either with feedback from others about their category membership, or (more probably) a mixture of trial and error learning and unsupervised observational learning. But there are also many concepts that are learned first through direct instruction and then incorporated into one's working conceptual repertoire through exercise of the concepts in real cases. Most forms of expertise – be it in finance, medicine or horticulture are likely to be learned in this way. Concepts (and thus the meaning of the words that label them) are learned through experience in different situations. Knowledge of their logical properties (such as the inferences they support) is stored side-by-side with knowledge of their perceptual/sensory qualities, their emotional valence and their potential for action and achieving goals (Barsalou, 2008). Understanding how nouns and verbs contribute meaning to utterances is likely to be dependent on a full treatment of the richness of our conceptual representations.

Conclusions

Psychology has generated a number of theoretical models for the concepts that underlie the meanings of nouns. Reviewing the large number of studies that have been done within the field, the conclusion one arrives at is that there is actually good evidence for each of them. In fact, different concept meanings may require different accounts of their representation. Some domains may be represented by linearly discriminable concepts with simple prototype structure. Others may have a *granularity* such that the internal structure of a category has sub-prototypes within it. A closer examination of the differences between prototype and exemplar theories suggests that they are simply different versions of the same model. Barsalou (1990) showed that the models are at either end of a continuum with maximum abstraction at the prototype end, and zero information loss at the exemplar end. Different concepts probably lie somewhere in between, with the degree of abstraction depending on the specific conceptual domain. Highly common and highly similar entities (like rain drops) may be represented as an abstract prototype, while familiar and distinctive classes, like the concept of dog for a dog lover would be represented in granular fashion as the collection of individual dog breeds, themselves represented perhaps by prototypes.

In addition to these similarity-based models, other concepts involve detailed schematic knowledge of the kind found in science and other academic disciplines, where similarity becomes less relevant and explicit definitions more common.

Given the variety and flexibility of the mind in how it provides meaning to the world, it should not be a surprise to find that our words have similar multiplicity in how their meanings are constructed in the mind. Words are used for many functions, and this necessarily gives rise to a wide range of semantic structures.

References

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. (2008). Cognitive and neural contributions to understanding the conceptual system. *Current Directions in Psychological Science*, *17,* 91-95.

Barsalou, L. W. & Sewell, D. R. (1984). Constructing representations of categories from different points of view. *(Rep.No.2).Emory University, Atlanta GA: Emory Cognition Project*.

Collins, A. M. & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407-428.

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

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel,W., Dry, M., Voorspoels, W., & Storms G. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods, 40 (4)*, 1030-1048

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

Hampton, J. A. (1981). An Investigation of the Nature of Abstract Concepts. *Memory & Cognition, 9,* 149-156.

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

Hampton, J. A. (1984). The Verification of Category and Property Statements. *Memory & Cognition, 12,* 345-354.

Hampton, J. A. (2011).Conceptual Combinations and Fuzzy Logic. In R.Belohlavek and G.J.Klir (Eds.) *Concepts and Fuzzy Logic,* (pp. 209-231). Cambridge: MIT Press.

Hampton, J.A. (2012). Thinking intuitively: The rich (and at times illogical) world of concepts. *Current Directions in Psychological Science*, *21,* 398-402*.*

Hampton, J. A., Dubois, D., & Yeh, W. (2006). The effects of pragmatic context on classification in natural categories. *Memory & Cognition, 34,* 1431-1443.

Heit, E. & Barsalou, L. W. (1996). The instantiation principle in natural categories. *Memory, 4,* 413-451.

Jönsson, M. L. & Hampton, J. A. (2006). The Inverse Conjunction Fallacy. *Journal of Memory and Language, 55,* 317-334.

Keefe, R. & Smith, P. (1997). Theories of vagueness. In R.Keefe & P. Smith (Eds.), *Vagueness: a reader* (pp. 1-57). Cambridge: MIT Press.

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

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

Nosofsky, R. M. (1988). Similarity, frequency and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54-65.

Osherson, D. N. & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition, 11,* 35-58.

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

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

Roth, E. M. & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology, 15,* 346-378.

Ruts, W., Storms, G., & Hampton, J. A. (2004). Linear Separability in Superordinate Natural Language Concepts. *Memory & Cognition, 32,* 83-95.

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.

Storms, G. (2004). Exemplar models in the study of natural language concepts. *Psychology of Learning and Motivation-Advances in Research and Theory, 42,* 1-40.

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

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

Further Reading

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577-609.

Hampton, J. A. (2006). Concepts as Prototypes. In B.H.Ross (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 46* (pp. 79-113). Amsterdam: Elsevier.

Hampton, J. A. (2013). Concepts in the Semantic Triangle. Chapter in E.Margolis and S.Laurence, *Concepts: New Directions.* Cambridge: MIT Press.

Hampton, J. A., & Jönsson, M.L (2013). Typicality and compositionality: The logic of combining vague concepts. In M.Werning, W.Hintzen, and E.Machery (Eds.), pp 385-402. *Oxford Handbook of Compositionality*. Oxford: Oxford University Press.

Murphy, G. L. (2002). *The Big Book of Concepts.* Cambridge: MIT Press.