CONCEPTS AS PROTOTYPES

James A. Hampton
City University, London

".. whatever vagueness is to be found in my words must be attributed to our ancestors for not having been predominantly interested in logic."

Bertrand Russell, 1923.

I. Introduction

The Prototype Theory of conceptual representation in large part owes its beginnings to Rosch and Mervis (1975), who, in the space of a couple of years, published a string of major papers laying out the empirical basis for the theory. The motivation for the theory came from a perceived crisis in philosophy and linguistics to do with defining the meaning of words. To the lay person, who has never worried too much about such things, the meaning of words is just given in the dictionary. The trouble is that most dictionary definitions are really only approximate or partial. The word "red" for example is not defined by a fixed interval of the color spectrum, but is the name for a imprecisely defined region with vague edges. The word "chair" could perhaps be defined as a movable object made for sitting on that stands on the floor, and has a back. However once again the actual use of the word tends in practice to allow for vagueness – designers continually create new objects for sitting on and new contexts in which to sit, so that it is often unclear whether they should be counted as chairs or

THE PSYCHOLOGY OF LEARNING
AND MOTIVATION VOL. 46, 79 - 113

not. The central insight of prototype theory is that word meanings, and the conceptual classes that the words name, are distinguished one from another not in terms of an explicit definition, but in terms of similarity to a generic or best example. The concept red is the class of colors that are centered around a particular point on the spectrum that everyone tends to agree is the *prototype* red. Indeed Berlin & Kay (1969) reported that there was better agreement about the *best* examples of color terms, than there was about the boundary between one color and another (for example between red and orange). The category of red things is therefore the category of things whose color is sufficiently similar to a prototypical red (and dissimilar from other prototypes). Similarly there are concept representations for "chair" and "stool" and "bench" and "sofa", each of which is associated with a prototype example of the class. Objects are then classified on the basis of which prototype they are most similar to.

Rosch, Simpson and Miller (1976) showed that people could readily learn novel categories based around prototypes (a point already demonstrated by Posner & Keele, 1968), and Rosch and Mervis (1975) analysed a number of semantic categories such as fruit, sport or vehicle to show that what members of the category had in common was not some set of defining features, but a sufficient degree of resemblance to each other. In some of their writings it is implied that the *best example* of the category, whatever that might be, would be the prototype. However it quickly became clear that the prototype should better be considered as a more abstract, generic concept, that was constituted from the different ways in which the category members resembled each other, and differed from non-members. Unlike a best example, an abstract prototype allows for the representation of different possible values of relevant features – such as that apples can be red, green, brown, or yellow, or that furniture can be sat on, slept on, used for storing things, or provide a surface for supporting things. An apple that had *all* these colors, or a piece of furniture that served *all* these functions would not necessarily be prototypical.

Prototypes then are the centers of clusters of similar objects, and prototype concepts form similarity-based categories. The center of the cluster is well established and agreed upon, but the boundary between one category and another may be subject to vagueness and disagreement. Talk of clusters with centers implies a spatial metaphor, and prototypes have often been discussed as points in *similarity space*. A mathematical exploration of the implications of this approach can be found in Gärdenfors (2000), and Osherson and Smith (1981) included a similarity space as part of their formalization of prototype theory. Spaces however have additional structural properties which impose unnecessarily strict constraints on prototypes. Verbeemen et al. (2004) have explored the degree to which natural categories can be represented in

spaces (through multidimensional scaling), and concluded that at least for some semantic domains, a non-spatial similarity model provides a better fit.

Following its introduction into cognitive psychology, prototype theory was also taken up enthusiastically by cognitive linguists such as Ross (1973) and Lakoff (1987), and anthropologists such as Kempton (1978) and Randall (1976). Ross (1973) for example proposed that the syntactic class NOUN in English is based around a prototype. He suggested a scale of "nouniness" associated with a hierarchy of syntactic acceptability in different contexts. The more nouny a word or phrase was, then the more contexts in which it would behave like a noun. A useful recent source of different views on the value of prototypes in linguistic theory can be found in Aarts et al. (2004).

While Rosch & Mervis provided overwhelming evidence for widespread prototype effects in semantic concepts and in category learning, the development of the theory in psychology subsequently remained relatively underspecified. In one of the last papers in the series, Eleanor Rosch (1978) discussed the theoretical underpinning of the data, and warned that a distinction should be made between the empirical phenomena of prototype effects, and any theoretical model that concepts are actually represented by prototypes. In fact, she doubted that the latter was the case.

The purpose of this paper will be to re-examine Prototype Theory and the evidence with which it is associated. One of the major difficulties with the theory may be that, with the early withdrawal of Rosch from the field, it has lacked a champion to develop and refine a working model of prototype representations, as new empirical results have been discovered. Thus, at various times, the theory has been criticised in many ways. For example, it is claimed that the theory lacks any way to represent the variability allowed on different dimensions within a category (for example the range of possible sizes of apples, rather than just their average size). The theory is said not to be able to account for some categories having wider or more flexible boundaries than others (and hence is unable to explain why a sphere half-way in size between a basketball and a watermelon is more likely to be a watermelon than a basketball, Rips, 1989). The theory is said to rely too heavily on statistical cue validity to determine feature weights (i.e. on the relative frequency of the feature for members and non-members of the category), and so to ignore causal dependencies among features such as that birds need their wings in order to fly. The theory is said to be circular in that no account is offered of why our attention is drawn to particular sets of features or particular sets of objects in the first place.

In every case, the criticisms may be well-founded, but what has been lacking is a coordinated attempt to modernise the theory to incorporate mechanisms to deal with the failures. It is of course easy to find data that a model has no way of explaining, if the model was not created with those data in mind. However one is then faced with a choice of discarding the model altogether, or of adapting the model to fit the data.

A notable exception to the lack of development of prototype theory has been the work on category learning of Don Homa and colleagues (e.g. Homa, 1984; Homa et al. 1981) and of J. David Smith and Paul Minda (e.g. Smith & Minda, 2000). Both of these groups of researchers have generated valuable evidence that in classification learning paradigms, there are conditions under which abstraction of prototypes does indeed occur. They have also developed precise quantitative models of how prototypes develop and are used in such learning situations. The question remains however whether the original aim of prototype theory – to provide an account of the natural concepts that we use to understand our everyday world and that serve to support the meanings of common nouns in natural language – can be met satisfactorily.

The chapter therefore will focus on the original evidence on which Prototype Theory was based, and will discuss which aspects of that evidence should be retained as central to the theory, and which aspects may be less crucial. I will also use this opportunity to present new results relating to prototype effects, and to reflect on some of the theoretical debates that surround the model. There is a nice irony here, in that the theory as applied to itself would suggest quite plausibly that "Prototype Theory" as a concept is itself a family of related concepts in which different importance might be attached to different assumptions of the theory. A prototype of Prototype Theory might be that presented by Rosch and Mervis, or that described in Hampton (1995), but other characterizations have been offered (e.g. Osherson & Smith, 1981). Leaving this irony aside, it is important first to try to capture the more essential characteristics of a prototype model, in order to consider how the central insights of the approach can be made consistent with recent evidence on the nature of conceptual representation.

II. The Origins of Prototype Theory

Prototype theory enjoyed rapid and considerable success in the years following Rosch and Mervis. Researchers were quick to apply the general notion of a prototype to a wide range of domains, such as clinical diagnosis and social stereotypes. The theory and its applications have been described in detail elsewhere (Hampton, 1997c; Murphy, 2002), so what follows will be a brief sketch. In general the way in which prototype structure was demonstrated for a domain was to establish one or more of four key phenomena about categories in that domain.

a) *Vagueness*. Categorization of items could be vague or "fuzzy". That is to say, there exist cases whose membership in a category is uncertain, not because of lack of knowledge, but because of the lack of a clear rule for categorization that applies to every case,
b) Typicality. Within a category, items differ reliably in their "goodness of example" or typicality,
c) *Genericity*. When asked to define or describe the meaning of a concept term, people tend to generate descriptions that are generically true of the class, though not true of all members,
d) *Opacity*. When asked for a rule that might be used to determine category membership, people are generally unable to come up with such a rule, and indeed even professional linguists typically find it an unrewarding goal (but see Wierzbicka, 1985). That is to say that the basis of categorization is not transparent to the speaker (as it would be when applying an explicit rule) – it is opaque.

These four phenomena are well-documented, and they constituted the basis on which domains as different as syntactical word classes, phonetic categories, speech perception, speech acts, psychiatric diagnosis, and personality perception were given a prototype treatment. For example in recognition of the fact that psychiatric disorders most typically have borderline cases, typical and atypical cases, symptoms that are commonly but not universally present in cases, and rely on a pattern of symptoms, rather than a set of necessary and sufficient criteria, the Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition (DSM-IV), published by the American Psychiatric Association in 1994 explicitly adopted prototype definitions for mental disorder categories. A person might be classifiable in a category providing that N out of M symptoms have been present for some given period of time.

The first two of the prototype phenomena relate to the "extensional structure" of a category – the way in which individuals and subclasses relate to the category. The second two relate to the "intension" of the category concept – the beliefs that a person holds that constitute the "narrow content" of the concept. Following Frege, *Descriptivist* theories of concepts, of which Prototype theory is an example, argue that the two kinds of information are intimately related. In some direct way, category extensions are determined through comparing an object or class with the intensional information constituting the concept. To decide what should be put in the category "bird" or "tool", any putative case is compared to the stored conceptual representation for that category, and a decision is computed based on the relation between what is known about the case and what is known about the category. Thus according to descriptivist theories, intensions determine extensions. Although many philosophers (e.g. Fodor, 1998; Millikan, 1998; Rey, 1983) have identified major difficulties with descriptivism, preferring to fix conceptual contents in terms of extensions (an *Externalist* theory of concept individuation), the large majority of cognitive psychologists still subscribe to this basic descriptivist position.

For Prototype Theory the determination of extension is achieved by specifying a measure of the match between the representation of an object or class and the prototype representing the category. If the degree of match is above some criterion, then the object is included in the category. If it is close to the criterion then it may be a borderline case, thus giving rise to Vagueness, and the further above criterion it is, the more typical a category member the item becomes, hence leading to the phenomenon of Typicality. The phenomena of Genericity and Opacity are found because a partial match to the prototype may yet be sufficient to be clearly included in the category – descriptive attributes that form a part of the representation may not be matched by all category members, and there may be no simple Boolean logical formula for devising a rule for categorization. (Note however that the similarity-to-prototype rule would correspond to a number of equivalent complex Boolean expressions, and that under certain circumstances it does in fact correspond to disjunctive or conjunctive rules – see Hampton, 1995).

Since these four phenomena are key to the proposal that a concept has prototype structure, the paper will proceed by taking each in turn, reviewing what is known about each phenomenon, and examining the validity of the arguments relating each effect to its supposed explanation. Alternative explanations offered by other theories of concepts (where they exist) will also be considered as appropriate.

III. Vagueness

A. Explanation Within Psychological Models

The existence of borderline cases for categories was first demonstrated systematically by McCloskey & Glucksberg (1978). They provided people with lists of items and asked them to classify them as members or non-members of a category. Their participants then repeated this exercise some weeks later. The finding was that for items that were independently rated as atypical of the category there was considerable disagreement about whether they should count as category members or not, and also considerable inconsistency across occasions, with people changing their minds as much as 25% of the time from one occasion to the next. It was also important to note that these items were not just unfamiliar. In a follow-up study, Hampton (1998)

showed that while a small number of McCloskey & Glucksberg's borderline items were indeed unfamiliar (is euglena an animal?), others were highly familiar (is a woman an animal?). With some few notable exceptions, Hampton (1998) found that for the majority of borderline cases, the probability of being placed in the category was directly predictable from judgments of how typical or representative the items were of the category prototype. It seems then that, at least in terms of empirical evidence, prototype theory is best placed to account for vagueness. Providing there is some randomness in the prototype representation or in the way that it is used, then we can expect probabilistic responding at the borderline, leading to disagreement and inconsistency in categorization.

Other psychological models can also explain vagueness. Exemplar models (Medin & Shaffer, 1978; Nosofsky, 1988) propose that concepts are similarity clusters very much like prototypes, with the exception that there is no central abstracted representation of the prototype. Instead there is a memory store containing a selection of actually encountered exemplars (items that have previously been categorized as falling under the concept). An item is categorized into the class to which it has maximum average similarity, the average being calculated across all stored exemplars. However categorization is explicitly probabilistic, with relative similarity to different categories determining the likelihood of being placed in one category rather than another. It is only once categories have been very well learned (or where the categories are very distinct from each other) that responding becomes all-or-none, and disagreement or inconsistency disappear.

The other class of models, theory-based or knowledge-based models, have in fact little to say about vagueness, but being Descriptivist, the same story can be given as for prototype and exemplar models. These models (Murphy & Medin, 1985; Rips, 1989; 2001) propose that concepts are individuated in terms of the role that they play in naïve theories that we use to explain our world. When this idea is cashed out into actual proposals for what is represented mentally, then concepts actually look very much like prototypes again, but with an important difference. Like a prototype representation there are different features or attributes involved; degree of membership depends on the features that a potential item may have; and typicality will depend on how closely an item resembles the paradigmatic case of a category member. The crucial difference is that similarity to prototype is not a simple function of matching attributes, but involves deeper causal information. One way to think of this is to suppose that in addition to having a set of features, a theory-based prototype has a set of information about the *relations* between those features. If an item has the features, but does not have them in the right relations to each other (which will include causal dependencies), then its similarity to the prototype will be poor. Authors of these theories of concepts may well resent the appropriation of their ideas into a form of prototype account, yet if they are to explain probabilistic responding, and residual effects of surface similarity in their data, (to say nothing of typicality effects) they are left with little alternative.

Taken more broadly, borderline cases are in fact an instance of a much more general problem – the problem of vagueness in natural language. Interest in vagueness goes back at least as far as the Ancient Greek philosopher Eubulides of Megara, who devised the Sorites Paradox to illustrate the problem. Sorites means a heap in Greek, and the paradox involves asking how many grains of sand are needed to constitute a heap of sand. It appears for example that removing a single grain from a heap could not of itself turn a heap into a non-heap. Yet repeating the action will eventually leave no sand left at all, so that at some point the heap must cease to be a heap. In fact this must presumably happen before the number of grains reaches some (again unspecified) small number. The paradox can be run in the opposite direction as well, by starting with a single grain (not a heap) and then asking if addition of one grain could turn the collection of grains into a heap.

B. Philosophical Accounts of Vagueness

Resolution of the problem of vagueness remains a current goal in philosophy, logic and indeed psychology (Hampton, 2005; Kamp & Partee, 1995; Keefe & Smith, 1997; Osherson & Smith, 1997). A notable attempt to resolve the issue is to relate vagueness to epistemological uncertainty. Williamson (1994) has developed an epistemological account of vagueness, in which it is claimed that the meaning of terms is actually precise, but that we all have different partial understanding of what that meaning may be. Because the true meaning may be highly complex, and does not correspond to any simple definitional rule, the average language user learns to approximate to that meaning. This approach is an example of the Externalist view of concepts/meaning, described above. A concept is something external to the thinker, that we come to represent in our minds more or less accurately as the case may be. Concepts and the meaning of terms that name them are constituted by the existence of a particular class in the external world. Our representation of that class may therefore show signs of inaccuracy and vagueness. Hence disagreement and inconsistency are to be expected, just as if one asked people to rank order a set of rivers in terms of their length, or historical events in terms of their chronological order. Different people will know the answer with different degrees of accuracy and reliability. It is of course also possible that the true meaning is not a class in the external world but a type of language use sanctioned by the social structure of the language group. In both cases

however, the definition of a term is external to the individual, and so vagueness could reflect a partial grasp of that definition.

This is clearly a defensible position if one takes an Externalist view of what a concept is – namely an entity that exists in external reality rather than in our heads. The position is also (perhaps paradoxically) quite consistent with the existence of prototypes in our minds – as Fodor (1998) has made quite clear. It may be the case that the concept of an X is clearly defined and delineated in the real world, but that my understanding of it is sufficiently partial that I am unable to decide clearly whether a particular sample is actually an X. My understanding of the concept could in fact be a descriptivist prototype, acquired from experience with typical cases of X, that has led me to form an internal representation of beliefs about Xs.

There is a sensible view, expressed by Bertrand Russell (1923) that says that vagueness is inherent in the relation between a representation and the world. There is no vagueness in the world itself. As he wrote: "things are what they are, and there is an end of it. Nothing is more or less what it is, or to a certain extent possessed of the properties which it possesses." The challenge for the Externalist view of concepts is then to find anything at all to say about the properties of concept classes, given that the very act of describing those properties introduces a symbolic representation which must on all accounts involve vagueness. In fact one would have liked to ask Russell what possible candidate "properties" he had in mind. It would be a trick question of course, since he could only answer in language in which by necessity the property in question would be vague, and so a thing might possess it to only a certain extent. Perhaps it is better to think of things being what they are, and not of "having properties" at all (see Quine, 1948, and Mellor & Oliver, 1997 for discussion of this possibility). Perhaps the very notion of ascribing a property to a thing is to create the logical problem of potential vagueness.

C. Further Studies of Vagueness

What do people say about the vagueness of their categories? In a recent study (Hampton, 2004) I presented people with 8 category lists that included borderline cases and non-members, and first asked them to rate each word with one of three responses "definitely in the category" "intermediate" and "definitely not in the category". Once people had given their ratings they were then asked to go through the booklet once more and indicate which of a number of possible reasons they might have for giving an "intermediate" response. The most common reasons chosen were variability of criterion ("because it depends on whether you take the category in a broad or in a narrow sense"), 31%, and epistemological uncertainty about the item ("because I don't know enough about the item to say"), 25%. Two other reasons that reached double figure percentages were category polysemy ("because it depends on how you define the category "), 15%, and item polysemy ("because it depends on how you define the item"), 11%. So it seems that people's intuitions about vagueness do include the possibility that they did not know enough about an item, but at the same time they see the category terms as being vague in the sense of having broader and narrower senses, and having different ways of being defined. Both of the latter reasons are consistent with prototype representations with variable dimensional weights and variable criterion placement. They are not however consistent with the idea that all vagueness is caused by uncertainty.

A follow-up to this study considered the question of stability of category decisions over time. One particular suggestion for handling vagueness in logic, supervaluation theory (Kamp & Partee, 1995), proposes that there is a given region of vagueness at the boundary of a category. It should therefore be the case that if people were given three response choices in categorization – "definitely yes", "possibly", and "definitely no", there should be less inconsistency in a test-retest measure. The idea is that people may not know how to categorize the "possible" items, and so may shift their decisions about this subset of vague borderline cases, responding in a probabilistic way depending on their current whim. Yet they may have a much clearer idea of what is definitely in the category and what is definitely not in the category.

The study, conducted with my student Bayo Aina, involved two groups of participants. One group categorized the same lists of category items using a traditional "yes"/"no" decision, while the second group had three options as outlined in the previous paragraph. Both groups returned a week later to make the decision again. The proportion of responses remaining the same on the second occasion was 83% in the two-choice condition, and only 73% in the three choice condition. So there was clearly no well-defined boundary region that people could easily discriminate. Why though was there a *drop* in consistency given the 3 options? One possibility is that with 3 options there are more opportunities to change your mind than with just 2. We therefore reanalysed the data, collapsing the 3 options into 2 by either comparing definitely yes with the other two, or comparing definitely no with the other two. Is it perhaps the case that we can consistently judge what is "definitely" in a category, but find it harder to judge what is definitely not? When the boundary between "definitely yes" and the other options was examined, 84% of responses remained the same, and exactly the same degree of stability was observed between "definitely no" and the other two options. The results therefore supported the view that instability due to vagueness is the same across the category scale. No matter whether the criterion is set high ("definitely yes" versus "not definitely yes"),

in the middle ("yes" versus "no") or low ("definitely not" versus "not definitely not"), the same degree of inconsistency in responding was observed. A final note about this study: over 5% of responses were changed from "definitely yes" to "definitely no" or vice versa, and only 1 of 28 participants failed to make any radical changes of this kind. Requiring a high level of confidence in a response was no protection against at least some degree of vagueness.

In the remaining part of this section, I turn in more detail to the question of how prototype theory actually accounts for the empirical behavioural phenomena of vagueness and instability in categorization. I will consider three of these – the disagreement between individuals on what should be included in a category, the instability of categorization judgments over time within the same individual, and the question of what type of logic can be applied to vague category propositions.

D. Accounting for the Data

1. Between-Person Disagreement in Categorization

The fact that disagreement between individuals is generally higher than the level of inconsistency within individuals (Barsalou, 1987) should probably be put down to the socially mediated nature of conceptual contents. Words in a language represent not a single meaning but a family of possible meanings that will naturally drift and evolve over time. Each individual keeps track of how others are using words, and so there is an approximate convergence in meaning within a given linguistic community. In this case, the observed probability of categorizing a word in a given category reflects the frequency of that belief/practice within the sample of people tested. Clearly, any theory of conceptual representation would need to subscribe to a similar account to explain the existence of disagreement. However it is important to note how readily the prototype representation allows for families of closely related meanings. In fact similarity-based categorization is the only theory that explains how variation in meaning appears to be *continuous* rather than discrete.

If concepts were based on definitions, or on theories of causal determination, it is likely that different meanings would be discrete, and hence more easily differentiated – as when we notice partial translation between languages (the French word "*fruit"* does not include nuts, whereas for many English speakers the equivalent English word "fruit" does.) There is no evidence that individual variation in conceptual representation shows clusters that might correspond to some small number of different meanings possessed by different groups of individuals (although to be fair few attempts have been made to look for such evidence). Having a prototype with continuously variable dimensional weights captures this non-discrete nature of conceptual variation very neatly.

2. Within Individual Inconsistency in Categorization

Within-individual inconsistency itself requires some external or random process to be at work. After all, any determinate system for using intensional information to perform categorization will always be perfectly consistent, whether it is based on a simple definitional rule or based on similarity computed across multiple dimensions. One suggestion for the source of the inconsistency (Braisby, 1993) has been that participants may recruit different contexts or perspectives in responding on each occasion. If a person's representation of a category varies according to factors such as the communicative context or the implicit contrast category, then failing to control such influences would contribute to inconsistency. There is however, currently, little or no evidence that within-individual inconsistency is due to failure to specify the context. At least one recent study failed to show any change in inconsistency when context was added. Hampton, Dubois & Yeh (in press) provided people with two strongly different purposes for providing a classification of common everyday category items. In one condition they were asked to classify pragmatically – placing items in the category where they thought that most people would expect to find them. In a different condition, they were asked to classify in a quasi-legal context – for example classifying artifacts as tools or furniture so that import tax regulations would be fairly applied, or classifying different academic pursuits as science or not so that a Science Funding agency would know how wide to draw the remit of their activities. Neither condition showed any reduction in individual inconsistency compared with a No-context control condition, and in fact all three conditions showed high levels of correlation between likelihood of categorizing an item and context-free judgements of its typicality in the category.

In the absence of evidence to the contrary, inconsistency is more likely to reflect random variation in processing. Barsalou (1987) suggested that recent experiences and random influences on memory retrieval probably lead to different conceptual representations being constructed in working memory on each occasion prior to the categorization. Requiring the intermediate stage of constructing the conceptual representation in working memory means of course that we have less direct evidence about the structure of the long-term store from which the information is retrieved. It is therefore difficult to judge to what extent the randomness occurs in the process of information retrieval, as opposed to in the permanent semantic memory store itself.

3. Prototypes and Logic

The final issue arising from category vagueness concerns the mapping of sentences in natural language onto logic. We commonly like to think that when we make assertions then the things that we say may be true or false. Nothing could be plainer. However it turns out that within every statement there is a sometimes uneasy trade-off between the truth or falsity of the statement and the interpretation of the words within it (Bill Clinton's narrow legal definition of "sex" in the Monica Lewinksi scandal is a good case in point). If there are borderline cases of category membership, then how does one handle the truth of statements that involve such cases? The problem lies in the famous dictum of the Law of the Excluded Middle. As Frege (1903/1970) put it:

> A concept that is not sharply defined is wrongly termed a concept. Such quasi-conceptual constructions cannot be recognized as concepts by logic; it is impossible to lay down precise laws for them. The law of excluded middle is really just another form of the requirement that the concept should have a sharp boundary…….Has the question 'Are we still Christians?' really got a sense, if it is indeterminate whom the predicate 'Christian' can truly be asserted of, and who must be refused it?

Early attempts to rescue the situation with Zadeh's fuzzy set logic (Zadeh, 1965) came to grief as it was quickly noticed that while (probably) a consistent logic in itself, with useful applications in control engineering, the logic made the wrong predictions about behavioral data such as judgments of typicality, or categorization in complex concepts (Osherson & Smith, 1981; 1982; Roth & Mervis, 1983). Hampton (1997b) reviewed a series of studies I conducted on this question from which it is clear that when people form conjunctions (Sports which are also Games), disjunctions (Fruits or Vegetables) or complement conjunctions (Dwellings which are not Buildings) they do not respect the constraints of set logic – fuzzy or otherwise. As a brief example, people say that chess is a sport which is a game, but that it is not a sport, they say that a mushroom is not a fruit and that it is not a vegetable, but that it *is* one or the other, and they say that a tent is not a dwelling, but that it *is* a dwelling that is not a building. These studies and others (e.g. Cohen & Murphy, 1984) strongly suggest that people form quasi-logical combinations of nouns using the natural language conjunctives "and" "or" and "not", not by forming Boolean set intersections, unions or complements, but by combining the *prototypes* of the concepts in question (see Hampton, 1987). As a further example, Hampton (1996) showed that judgments of membership in a conjunction showed *compensation*. The more typical an item was as a member of class A, then the less similar it needed to be to B, to be counted as a member of the conjunction A^B. For example if judging whether faces are those of a "happy child", the more typically childish an already clearly childish face became, the less happy the child needed to look in order to still count in the category. Logical conjunction just doesn't work this way.

The discovery of this non-logical system for combining concepts is one of the key factors supporting prototype representations, since it flies in the face of grounding the meaning of terms in extensionally delineated classes in the world, and of grounding complex concepts in set logic. It is not then surprising to find the whole process of conceptual combination becoming one of the major battlegrounds in the debate between externalist and descriptivist theories of concepts. In a number of books and papers, Fodor (e.g. Fodor & Lepore, 1994) has presented the case that concepts cannot be prototypes as follows (I paraphrase):

a. It is a fundamental tenet of the representational theory of mind that thought is compositional – that is that the meaning of a complex thought is solely made up from the meaning of its component parts, and the syntactical function of the linguistic structure that links them together.
b. Concepts are the component parts from which complex thoughts are created
c. Therefore concepts must compose in the way stated
d. Prototypes do not compose in this way, therefore concepts cannot be prototypes.

Fodor doesn't claim that concepts don't *have* prototypes, just that they are not themselves prototypes. So, terminological tussles aside, what we have is the suggestion that the entities that psychologists study and like to call concepts, are not in fact concepts, and might better be called conceptions or prototypes. A different level of mental representation contains our concepts. These concepts are atomistic symbols (cannot be further analysed into simpler terms such as descriptions), and have the requisite properties of composing according to Boolean logic. It is possession of these concepts that explains the compositional properties of our thought.

It will be interesting to see if this proposal can be cashed out into empirical predictions about those circumstances in which concepts "proper" are involved in thinking, and those in which we rely instead on our prototypes. One possible way forward may be in differentiating rule-based and similarity-based systems of thinking (Smith & Sloman, 1994). Ashby et al. (1998) have intriguing data that in category learning there are two independent systems that learn through either hypothesis testing of rules or through accumulation of associative similarity-based links, and that these are associated with different brain regions. There is also assorted evidence emerging that individuals differ systematically in whether they use similarity or rules in solving conceptual problems (Hampton & Estes, 2000; Winman et al., 2005).

The issue of how to marry our ability to think in logical terms with the flexibility and adaptability of our conceptual system is a key issue for cognitive science. Clearly if we went around thinking that something could be an A which is a B, but at the same time not an A, we would be continually falling into reasoning errors. In fact, when faced with logical arguments dressed up in real world situations, it would appear that most people find it very hard to judge the logical validity of arguments (Henle & Michael, 1956). My guess is that thinking in terms of set logic and compositional concepts is a relatively late cultural acquisition that arose with the development of civilisations involving technology, economic accounting and mathematics in the last few thousand years. To use language for logical thinking requires that we *stipulate* and then hold constant the meaning of words in the given context, so that Frege's dictum of sharp boundaries can be respected. For example, we could answer Frege's question, "Are we still Christians?" in one of two ways. Following Frege we could stipulate (for example) that a Christian is one who is baptised into some closed set of recognized churches– hence everyone on the planet is either a Christian or not a Christian. If we then stipulate who Frege is referring to by"we" we can check whether the set defined by "we" is included within the set defined by "Christian". QED. Alternatively we can take the question in a non-logical way, as asking (perhaps even rhetorically) whether the current trends in our religious beliefs and practices have taken us away from the original "true" notion of Christianity. This way of answering the question requires a discussion about the true meaning of Christianity – it becomes no longer a question about sets of entities and their set relations, but a question about concepts and how they should be defined. Having identified the fundamental core beliefs and values that we want the term "Christian" to imply, we can then come to some broad judgment about the degree to which such values are prevalent within the group of people defined by "we" (which in turn may not be well-defined as a group, but admit of clear and borderline members).

This example illustrates the problem that we have. When we use language we may be either referring to sets in the world or alternatively asserting the meaning of our words. I suspect that sentences such as "the cat is on the mat" with literal interpretations and straightforward truth evaluation, are quite rare in our daily discourse.

4. Conclusions about Vagueness

In sum, category vagueness provides support for prototype representations given an additional assumption that the representation itself or the processes that utilize that representation are subject to random or contextual noise. It is interesting that other psychological accounts of concepts, such as the theory-based view (Murphy & Medin, 1985) have little to contribute to the discussion about vagueness. It seems that disagreement and inconsistency of categorization of familiar items, and the close link between probability of categorization and similarity to the category are key pieces of evidence in favour of the prototype view. Exemplar models however make much the same predictions, since they share with prototypes the idea that categorization is similarity-based and probabilistic.

Alternative philosophical accounts of the phenomenon of vagueness exist which do not require that concepts be prototypes. In the case of epistemological uncertainty however, it may be possible for peaceful coexistence between an externalist account of concepts and a prototype-based account of our mental representations of those concepts (see also Prinz, 2002). When it comes to the use of concepts as elements of thoughts, to be combined compositionally through logical operations, then prototypes do not have the right properties. They have been shown to combine in non-logical ways, and do not respect the clean rules of set logic. To some, this is devastating news for prototypes as a component of the future of cognitive science. Our conceptual thought is logical, so our thought cannot be based on prototypes. To others, this fact about prototypes goes some way to explaining the vast literature on human reasoning – we just are very bad at thinking logically most of the time, whereas we are pretty good at shifting the meaning of our terms mid-argument if it will suit our purposes.

IV. Typicality

Variation in the typicality of category members is often cited as one of the core tenets of prototype theory. However it is questionable whether the simple fact of typicality variation itself is particularly discriminating between prototype theory and other accounts of concepts. The problem is that when instructed to judge typicality or goodness-of-example it may be unclear just what aspect of the category members people may be attending to. Barsalou (1985) found that there were several different factors involved in determining mean typicality scores for common taxonomic categories like Bird or Fruit, including resemblance to other category members (as predicted by prototype theory) but also frequency of instantiation (how often the item is encountered) and fit to ideals (how well the item meets some goal or purpose – for example for artifact concepts). Subsequently Medin and Atran (2004) have reported that in non-student populations the notion of "goodness of example" as originally

introduced by Rosch is much more likely to be determined by fitness to goals or ideals than by similarity to other category members. The point was driven home by Armstrong, Gleitman and Gleitman (1983) who showed that participants were just as happy to rate the typicality of well-defined categories such as "odd-number" as they were to judge typicality in allegedly prototype categories such as "fruit" (although see Larochelle et al. (2000) for evidence about how typicality effects in well-defined categories differ from those in other kinds of category).

The moral would seem to be that given a task in which category members have to be ranked or rated for goodness-of-example, people will quite reasonably attend to whatever dimension is available that provides sufficient information to rank the items. For natural kinds this may include similarity to prototype and frequency of instantiation, for artifacts it may include ability to meet the functional goals of the artefact, and for mathematical concepts it may reflect familiarity, simplicity or yet other dimensions. In fact there may be very few open-ended categories that do not have reliably measurable typicality differences within them. After all John is a more typical name for a British male than is Tyrone, but this is presumably not because it is more similar to other names.

So the demonstration of reliable typicality differences may be neither here nor there – prototype concepts would certainly be expected to show such differences, but then so might other kinds of concepts given the ambiguity of the task. It is therefore important to go beyond the ratings themselves and examine what other behavioral effects can be associated with typicality differences. Here the results suggest that variations in typicality proper (i.e. rather than frequency or familiarity) have strong and robust effects on a range of psychological tasks, consistent with similarity-based models (which would include Exemplar models). It is therefore encumbent on theories of concepts that have little or nothing to say about typicality to provide an alternative explanation of these effects. In the following section I review some of these results.

A. Some typicality effects

One of the first demonstrations of typicality effects was the finding that typical category members are more rapidly categorized than atypical (Rips, Shoben & Smith, 1973; McCloskey & Glucksberg, 1979). But is this effect due to degree of similarity to the prototype, or due to associative strength between the member and its category? Hampton (1997a) explored two dimensions of category gradedness and their effects on the speed of categorization of category members. Using the British category norms published by Hampton & Gardiner (1983), two sets of materials were constructed. One set provided a contrast between words that were of high typicality and others of low typicality, while holding constant their production frequency in the category norms. For example, for the category Birds, typical items *nightingale, swift* and *dove* were compared with atypical items *ostrich, penguin* and *emu.* Both kinds of word were equally likely to be generated as exemplars of the category, and were rated as equally familiar, but one set was judged as typical and the other set as atypical. The second set of materials arranged the converse comparison – a contrast between words of high production frequency (PF) and words of low production frequency, with the rated typicality held constant. For example high PF birds *eagle*, *hawk* and *duck*, were contrasted with low PF *cuckoo, peacock* and *turkey*. These sets were matched for typicality and familiarity, but the high PF words were generated frequently when listing category members, and the low PF words were not.

These two different contrasts were measured under two experimental manipulations. One factor (adapted from McCloskey &Glucksberg, 1979) varied the difficulty of the categorization task by manipulating whether the false items, to be rejected, were obviously false and unrelated to the category (bus – bird) or whether they included related items that would bear some similarity to the category (bat – bird). In one condition therefore false items could be easily rejected, and the instructions emphasized speed of response, whereas in the other condition false items were confusable, and the instructions emphasized accuracy of the response. The second factor that was manipulated was priming of the availability of half of the items by providing a prior task in which they were categorized at a more superordinate level (e.g. sparrow – animal).

The purpose of these two manipulations was to show that typicality and associative strength (PF) could be doubly dissociated in the context of speeded categorization of words. False relatedness was intended to make the atypical items harder to categorize, and so enhance the typicality effect, whereas priming was expected to temporarily increase the accessibility of all items, and so reduce the PF effect, which relies on the inaccessibility of low PF items.

The results showed that introducing related false items did increase the effect of typicality on response time, but did not interact with production frequency. Alternatively, priming decisions with a superordinate category reduced the production frequency effect in the easy condition, but did not interact with typicality. Error data confirmed the picture, with false relatedness and accuracy instructions increasing "no" responses to atypical items, but reducing "no" responses to low production frequency items.

Data such as these provide strong support for the general notion that categorization of items in common semantic categories can involve a similarity-based comparison of

the item with some generic representation of the category. Increasing the difficulty of discriminating the false items from the true items requires that a greater amount of information needs to be retrieved, with the result that a greater difference is seen in response time for typical vs. atypical items, and atypical items are more likely to be rejected from the category. Associative links between items and the category name provide a separate and dissociable source of variance between items affecting categorization time. Hence typicality is not just associative strength.

Typicality continues to prove itself an important variable. A recent study by Kiran and Thompson (2003) will serve to illustrate this. They set out to treat naming deficits in four patients with fluent aphasia. Over many weeks the patients were trained in category sorting and naming of pictures, identifying semantic attributes applicable to target pictures and answering yes/no questions about the features of the target. Patients were either trained with a set of 8 typical category items, or with a set of 8 atypical items, and generalisation was tested to 16 other category members. The results were striking. Training on atypical items generalized to the rest of the category, whereas training on typical items did not. If one conceives of the category concept as being represented by a prototype in a feature space then clearly activation of widely spaced atypical examples will generalize to the whole region of the space, whereas activation of a cluster of typical examples near the centre will generalize less widely. (Similar conclusions were drawn from a quite different paradigm – the release from Proactive Interference in short term recall – by Keller & Kellas, 1978).

B. Stability of typicality judgments

As with categorization decisions, there is also considerable variability in people's typicality judgments. Barsalou (1987) conducted a series of studies of the instability of typicality ratings and rankings, and concluded that the high levels of shift in an individual's ratings from one occasion to another argued for prototypes being constructed in working memory anew each time a typicality rating task was presented. In a recent study conducted with my student Lara Olufon we set out further to investigate the within-participant stability of typicality ratings. In particular we tested a prediction of prototype theory that had not been tested before. One plausible source of variability in typicality judgments would be variation in the relative weight given to different aspects of the prototype. Perhaps on one occasion a person feels that being sweet is the most important feature of a typical fruit, whereas on other occasions they feel that being round is more important. The effect of this variation will be that the relative similarity of items to the prototype will change. However this change will only be observed for items that have one but not the other of the features. Items that have all of the features will still be the most typical, regardless of any shift in weight from one feature to another. In spatial terms, shifts in dimensional weight that stretch or shrink different dimensions will leave the centre of the category unmoved, although distance of atypical items from the centre will be affected. We therefore predicted that the items judged most typical would be least likely to shift their ranking on a retest a week later.

Note that it is also true that items with few or none of the prototype features should show less variability. However lacking enough features, these items would not fall in the category, and so would not be included in a list of category members. We hypothesized that all items in the category would have at least half of the full set of weighted features. Hence variability should increase monotonically across the typicality scale within the category.

A possible confound here is the extra stability of items at the two ends of a sequence. For example, the item judged most (or least) typical will still tend to be most typical if its typicality increases, and will only risk a change in rank if its typicality decreases. So the chances of a change are half as great for an end item as they are for an item in the middle of the ranking order where a change either up or down on the scale risks a change in the rank position. Items in the middle of a rank order are also more likely to be jumped over by items on each side, than items at the end that have items on one side only. To control this confound, we compared the stability of items at the top end of the ranking (most typical), with those at the bottom end (least typical). Confounds due to position in the list relative to the end and the middle should be equal for both ends, so the predicted extra stability of typical items should show up as greater stability for the top end of the list compared with the bottom end of the list.

Nine category members were selected in each of 8 common taxonomic categories studied by Hampton & Gardiner (1983), such as birds, clothing and weapons. Care was taken to space the items equally along the typicality scale. In addition, 9 category features were selected for each category. Participants ranked the items for typicality and the features for their importance. They did the same task on two occasions a week apart, and correlations were calculated between the ranks given on each occasion. Median correlation between the first and second rankings was 0.77 for both feature importance and typicality rankings. Results also clearly showed greater stability for the top four ranks in the list than the bottom four ranks. Mean probability of the top four most typically ranked items retaining the same typicality rank was 0.33, and for the bottom four atypically ranked items was only 0.27. The difference

was significant on an ANOVA with end (top or bottom) and distance from the end (1 – 4) as within-subjects factors.

Consistent with the hypothesis that instability in people's concepts may reflect changes in the weights attributed to different prototype features, we therefore found that typical category members were more consistently ranked than were atypical category members.

As a final demonstration that typicality is an effect of similarity, rather than availability or some other variable, I conducted a short study with Wenchi Yeh in which participants gave typicality judgments for items that (unknown to them) were constructed in quadruples. Within a quadruple were two pairs of similar items, which when re-paired within the quadruple would constitute two pairs of dissimilar items. For example the pairs "goose and turkey" and "pelican and toucan" were similar pairs, which were then re-paired as "goose and pelican" or "turkey and toucan" to create dissimilar pairs. The measure taken was very simple – the degree of correlation in the ratings or rankings given to each member of a pair, across the different participants in the experiment. Thus for example, one group of students rated all the items for typicality, and then the correlation was calculated across individuals of the ratings given to each member of a similar pair such as *goose:turkey* and to each member of a matched dissimilar pair such as *goose:pelican*. The idea was that if individuals vary in the weight that they attach to different dimensions of a prototype, then there should be a stronger correlation for similar pairs than for dissimilar pairs. Having the same feature profile, similar pairs would move up or down together as feature weights changed across individual raters, whereas dissimilar pairs would not. This was what was found. The stronger the similarity between a pair of items, the larger the correlation between the ratings given to the items by different people.

C. Conclusions about typicality

In sum, typicality effects can be identified that are not simply to do with the familiarity or availability of category members. Theories of concepts that do not base categorization on similarity tend to be dismissive of typicality effects. Armstrong et al.'s (1983) results are often cited as discrediting the use of typicality to argue for prototype representations. However it is increasingly clear that a great many tasks are influenced by typicality effects, and these effects are rooted in differences of similarity or degree of match between an item and a conceptual representation. For prototype and exemplar theory, similarity-based typicality effects are a central plank of the models. The wide range of such effects is therefore a key piece of evidence for this type of theory.

V. Genericity

Genericity in linguistics and the philosophy of language refers to sentences that either (a) refer to a kind rather than a particular, or (b) assert general properties *typically* true of a class or individual. The following sentences illustrate these two phenomena:

> The potato was first cultivated in South America
> John smokes a cigar after dinner
> (examples and definition from Krifka et al. 1995, p2.)

In the first sentence "the potato" refers to the kind, not to an individual potato, whereas in the second the sentence implies that this is John's usual habit, and not that John never has dinner without smoking a cigar afterwards.

The two kinds of genericity coincide when people are asked to give general properties of a kind – which is the task that Rosch and Mervis (1975) used to develop prototype theory. For our purposes then, genericity refers to the finding that people generate descriptions that are typically true of the concept, where "typically true" implies that typical category members will have the property, but atypical category members may not.

A commonly observed phenomenon in all natural languages is the fact that many sentences may be neither universally true, nor simply false, but may instead be true under some notion of "generally true" or "typically true". When asked to describe birds and say what is distinctive about them compared with other related categories, people will commonly start with "has wings" and "flies", and then go on to describe other distinguishing features such as 2 legs, feathers, hatched from eggs and migratory. There appears to be no intuitive difference to the respondent between the relevance of saying that birds fly and saying that they have feathers. This in spite of the fact that there are well-known examples of flightless birds, and many species of insect that fly, whereas all birds (at least before they are prepared for the oven) and only birds have feathers. Given that there is a single defining feature – feathers – that is both necessary and sufficient to discriminate birds from other types of creature, why do people not recognize this fact and define the word's meaning in this simple way?

In fact, birds turn out to be a rather special and potentially misleading case. Early theories of semantic memory such as the classic paper by Smith, Shoben & Rips

(1974) took the fact that birds have a clear set of defining features, and other features that merely characterize typical birds as a basis for proposing a distinction between Defining and Characteristic Features, in spite of difficulties in establishing such a distinction for a broader range of concepts (Hampton, 1979; 1981; McNamara & Sternberg, 1983).

People's knowledge of what distinguishes other types of creature such as fish, insects, reptiles, or amphibians is much less easily captured in terms of a small set of necessary features. It appears that we apply a general approach to representing most of the conceptual classes that we distinguish in the external world – that is to represent them in terms of a set of information about what they typically are like, where they can be expected to be found, how they typically behave, or what humans typically use them for. Russell may have been right that "things are what they are", but our way of grasping reality involves setting up classifications that groups things together into classes or types – which almost necessarily will involve fuzziness in the categorization and the involvement of information that is distinctive and very useful to know, even though it lacks logical rigour.

Genericity is important evidence for the most central tenet of prototype theory (and more generally of similarity-based theories including Exemplar theory). This tenet is that:

Concepts are represented by their most common and distinctive attributes - if using a spatial metaphor, then they are represented as a region in semantic similarity space centred on a particular point corresponding to the most typical potential example of the category.

A corollary of this proposal is that the category boundary between one sort of thing and the next is not directly represented. We easily identify things when they are a good match to our stored representation, but we do not find it easy to draw fine distinctions at the category boundary. This proposal contrasts, for example, with rule-based categorization, in which it is the rule determining membership (and hence the boundary) that is represented, and not the collection of typical or usual features normally found in the category.

Genericity has been little challenged within psychology, although a number of different approaches to the problem of how to characterize the meaning of generic sentences have been proposed within philosophical semantics (see papers in Carlson & Pelletier, 1995).

A. GENERICITY EFFECTS

Genericity applies to most (but not all) statements of conceptual knowledge. Penguins live in cold climates (except for those in the Los Angeles zoo), tigers have stripes (except those born as albinos) and trees have leaves (except for deciduous trees in winter). One exception to the rule would be very general properties. Thus "penguins hatch from eggs", "tigers need food and water to live", are almost certainly universally true. (Or maybe it just gets harder and harder to imagine the scenario that would falsify them). In an early study (Hampton, 1982) I showed that even category membership statements need not be universally quantified. It might appear that to say "X is of type Y" is to mean that all Xs are Ys. However this constraint does not apply to all of our superordinate categories. People will say that a car headlight is a type of lamp, and that a lamp is a type of furniture, but the headlight is clearly not furniture. It is possible that this intransitivity is only observed with certain types of superordination – for example it does not occur readily with biological taxonomies. This point is made by Wierzbicka (1984), who argues that categories like furniture are really collections not types – in the way that tableware is stuff you put on your table, so furniture is just stuff that you use to furnish your house. However it is still puzzling why collections should not obey transitivity in the way we expect true classes to.

Some recent research I conducted with Martin Jönsson of Lund University, Sweden, examined the degree to which people's acceptance of the truth of generic statements changes as the subject noun is modified (Hampton & Jönsson, 2005). Taking our inspiration from a paper by Connolly et al. (2003), we considered sentences of the following type:

Ravens are black
Jungle ravens are black
Young jungle ravens are black

In keeping with Connolly et al.'s reported results, we found that as more modifiers are added to the subject noun, so the degree to which people think the sentence is likely to be true decreases. Connolly et al. interpreted this result as showing that people do not use the default prototype information for a general class (raven) when judging a subordinate (jungle raven), and as therefore undermining the notion that prototypes are involved in forming complex concepts such as jungle ravens. Our alternative account is that, following published models of how prototypes combine in intersective concept combination (Hampton, 1987; 1997b), one would expect there to be a reduction in the importance or weight of features that are true of one concept but not of the other. Thus if jungle creatures are not generally black, the intersection of ravens and jungle creatures will have less weight attached to that feature.

We thought it would be interesting to see how people responded when the sentences were universally quantified. Suppose now that you are asked to judge the likelihood that the following statements are true:

All ravens are black
All jungle ravens are black
All young jungle ravens are black

If people use their prototypes for constructing complex concepts in order to make these judgments, then they should continue to say that the sentences are less likely to be true as the number of modifiers increases. However if the presence of a universal quantifier triggers Fodorian atomistic concepts and logical intersection, then it cannot be the case that the first statement is more true than the others. Clearly in all worlds in which it is true that "all ravens are black", it is also true that "all jungle ravens are black" and similarly that "all young jungle ravens are black". If a property holds true of a whole class it must necessarily be true of any arbitrarily defined subset of that class.

Our results were overwhelmingly in favour of the prototype theory. Across both individuals and items the large majority had lower estimates of truth likelihood for the modified sentences than for the unmodified. We refer to this as the Inverse Conjunction Fallacy, since it takes the opposite form of Tversky and Kahneman's Conjunction Fallacy (Tversky & Kahneman, 1983), where likelihood estimates are greater for a conjunction of facts than for a single component fact.

We followed this experiment with one in which we varied the mutability of the attributes in the predicate part of each sentence. Mutability has been established by Sloman et al. (1998) as an important variable within conceptual representations. As discussed in an earlier section, certain attributes in a conceptual representation are involved in many causal or other dependency relations with other attributes – for example the motor of a car is involved in causal relations with the car's function, its need of fuel, its contribution to pollution etc. Such attributes tend to be less mutable – it is harder to imagine an example of the concept that is like other examples in every respect except missing just this one feature. We gave people a task in which they had to choose the more likely of two generic sentences with modified subjects, one with a mutable predicate (e.g. "Brazilian doves are white") and one with a less mutable predicate (e.g. "Brazilian doves have wings"). We discovered a strong preference for the sentence expressing the less mutable feature. Thus not only is attribute information inherited by the complex concept (Brazilian dove) from the simple noun concept prototype (dove), but the degree of confidence with which it is inherited depends on the internal structure of the prototype, in keeping with Hampton's (1987) account of the formation of Composite Prototypes.

B. Conclusions about genericity

Genericity is crucially important in the argument for prototypes. If it is true that we represent a concept in terms of its typical features, then there is no requirement that those features will be true of all members of the category, and indeed people may not even be aware, without conducting a memory search, of which features are universally true and which are not. Exemplar models would also not expect all features to be true of all category members, but here an important failing of exemplar models comes to the fore. Since they have been developed almost exclusively with respect to the categorization of individual particulars, there is very little that the models have to say about categorization of whole classes or kinds, or the truth of generic statements about a class. While they do a good job of predicting the learning dynamics and generalization performance for certain kinds of category structure, they have not been set the task of deciding whether a class as a whole has a particular property, or whether a class as a whole belongs in a superordinate class. In effect, the development of the models has been too restricted to tightly controlled artificial stimulus sets to offer much help with understanding many of the effects observed in natural language.

VI. Opacity: The Failure of Category Definitions

The fourth phenomenon considered to support a prototype view of concepts is the difficulty that has been encountered in generating good accurate definitions of the meanings of content words (particularly nouns and verbs) in any language. This problem was famously expounded in Wittgenstein's *Philosophical Investigations* (1953) in relation to the category of games. It appears that (in keeping with the discussion of genericity above), people know lots of things about games – they involve people, they take place over a period of time, they are done for their own sake, they involve rules, they involve winning and losing, they are unpredictable – but no set of these different features can be found that discriminates games from non-games, except by using a prototype rule.

In work originally done for my PhD (Hampton, 1979), I interviewed people about their definitions of 8 different semantic categories. The questions included asking what was true of all category members, what was true of typical members, what would

make something a borderline case – what features it would have, and what it would lack, and even what the word might mean if applied metaphorically to a thing or a person. Features were then listed, regardless of where in the interviews they were generated, and a separate group of people judged whether each of a list of potential category members had those features. Finally a third group categorized the list of items and made judgments about how good a member they were, or how related a non-member they were. The question was whether the set of category members could be distinguished in terms of a set of common features. The results were that half of the categories could be defined in this way, but half could not. A similar proportion of definable categories was found by McNamara & Sternberg (1983) using a procedure where each individual's definitions was compared to their own category judgments.

Of course the procedure is perhaps unnecessarily restrictive in its insistence on relying only on empirical evidence generated by respondents. Semanticists certainly take a much more unconstrained view of how the task should be done. Thus:

> ... semanticists are not obliged to take informants' judgments at face value.
>
> (Wierzbicka, 1990),

or

> ... why should the "real meaning" of a word correspond to what people think of as the meaning of that word? Folk theories should no more be a criterion in semantics than they are in syntax or any other aspect of linguistics.
>
> (Bouchard, 1995).

There is a serious issue here that arises frequently at the interface between psychology and other branches of cognitive science such as linguistics or philosophy. In a way reminiscent of the externalist theory of concepts propounded by philosophers, it is common for semanticists to see the analysis of word meaning as being the analysis of an abstract cultural artifact, such that a word's "true meaning in English" need not correspond directly to its current usage. This is a knotty problem that will take some unravelling. On the one hand, psychological methods can be accused with some justification of being crude and open to unwanted demand characteristics. We know, for example, that when generating features of a word's meaning people are driven by pragmatic considerations of trying to be as relevant as possible. Thus they may neglect to mention many features of birds (such as "has a heart") that they would nonetheless agree to be true, simply because they are less relevant to the perceived communicative goal of distinguishing birds from other creatures. More recently psychologists have also taken to asking people for metalinguistic reflections on concept meaning – for example rather than asking for the definition of a term, asking whether the term has a definition (Armstrong et al, 1983), or asking whether membership in a class is all-or-none or graded (Estes, 2004; Kalish, 1995). This methodology, while instructive, is also subject to the same question – at what level should we take what people say about how their mind works as a constraint on our theory of the mind? We don't study perception or attention this way (although researchers may get some useful ideas via introspection), but somehow one feels that conceptual contents just *are* what people claim they are. One is reminded of the wag who claimed that Wagner's music is not nearly as bad as it sounds.

There are some writers who still hold that given proper attention to the task and a degree of training, definitions of word meanings can be provided. Sutcliffe (1993) suggested that in following Wittgenstein psychologists have been looking in the wrong place for monothetic definitions of classes – it is not the many ways in which games resemble each other or differ that will tell you what constitutes a game, it is at the more abstract level in which games are differentiated from other broad classes of human activity. Wierzbicka (1972; 1985; 1987) has been the most tireless proponent of getting on with the task of actually giving definitions. For example her answer to Wittgenstein is as follows (Wierzbicka, 1990, p. 469):

Games

1. Things that people do
2. When they do something for some time
3. For pleasure
4. Imagining that they are in a world
5. Where they want to cause some things to happen
6. Where they know what they can do and what they cannot do
7. And where no one knows all that will happen

This definition is proposed to apply to board games, card games, ball games etc. and to exclude non-games such as a child idly throwing a ball against a wall and catching it again, which according to Wierzbicka would not be called a game in English. It is, in my view, a great pity that such definitions are not put to the test against a panel of competent speakers of English, rather than being tested against the author's (albeit expert) intuitions. It is easy to suggest potential counterexamples – category members that the definition excludes such as games that are played for money rather than pleasure (poker, professional golf), or games that are entirely predictable (simple computer games like space invaders), and non-members that are included such as watching Reality-TV shows and voting for one's least or most favourite participant in

the show, or standing on the touchline and shouting your support for your sports team. You may be influencing the outcome, you know what you are and are not allowed to do, but watching the game is not playing it.

The general ease with which definitions can be discredited explains the tendency of less brave individuals (in which group I include myself) to discuss the problem in general terms, without actually doing the descriptive work of proposing definitions for any particular term. A dedicated prototype theorist should be willing to take on the task of generating prototype representations, complete with feature weights, dependencies between features, and parameters for individual and group variability that would completely fit a range of data of the kind described in this paper for some set of concept terms. To date, only fragments of this project have been attempted, and probably with good reason. In the mean time, the prototype theorist takes some (unearned) comfort in the general failure of definitional approaches to word meaning.

For other psychological theories of concepts (other than exemplar theory), opacity may possibly be an embarrassment. If concepts derive their meaning from their role in a naïve theory, then what is to stop individual respondents from explaining that theory and hence providing the appropriate account of the concept. Theory-based models seem to be even more remiss than prototype models in making general claims about concepts, without providing a description of the actual contents of a concept. Their case would be much stronger if they could find a way to elicit the naïve theory for one or more domains from their respondents and then show how the theory affects decisions about the concepts within it. The experience of those interviewing experts in the course of developing Knowledge-Based Systems in Artificial Intelligence is that eliciting people's theories of a domain in an explicit format is a very difficult and time-consuming process.

In responding to the problem of opacity, and not favouring the prototype approach, philosophers have turned to two alternative accounts of word meaning, which for completeness sake I will briefly outline.

A. Deference

For deference, a famous paper by Putnam (1975) put the case in terms of a linguistic division of labour. Just as people accept that it is quite possible that they may be using a word in the wrong sense, and would refer to a dictionary to check – they also accept that for many terms referring to natural kinds such as Elephant or Gold there may be expert biologists with DNA testing kits or expert chemists with chemical assay kits who know the crucial tests of category membership. It is reasonable in such cases for lay speakers to defer to the relevant experts. As a result, in Putnam's memorable phrase "Cut the pie any way you like, 'meanings' just ain't in the head!".

Studies by Braisby (2005) suggest that people are indeed willing to change their minds about a categorization if an expert opinion differs from their own. However he found that there are quite restricted circumstances in which this will work, and in fact there is a considerable minority of individuals who strongly resist any attempt to influence their decisions by reference to experts. The notion of deference may therefore not offer a complete account of opacity. It also leaves us with two interesting questions: first how do the experts solve the problem of defining the concept, and second what happens for all those concepts for which there is no recognized body of experts?

B. Atomism

Probably the best way to define an elephant is to say that it is the result of two elephants breeding. This definition captures all and only elephants. It is not a truism, since it captures the commonly held belief that natural kinds have a particular set of features because of some germ-like essence which in the case of biological kinds gets passed from generation to generation. It even covers the concepts of more sophisticated individuals with a knowledge of evolutionary biology, for whom the concept of elephant would be vague at some point in the past as the species evolved from its predecessor.

The difficulty with this definition is two-fold. First, how do you know that two creatures who are about to have an offspring are indeed elephants? You seem to have doubled the problem you started with. And second, how do you know what kind of creature the parents of the creature now in front of you were? However note that these are problems of how we could come to know for sure what something is, rather than problems of not having the right definition. Fodor's atomist view is (roughly, see Fodor (2000)) that our minds contain an atomic symbol for the concept elephant. Through exposure to a world in which there are such things as elephants and, presumably, cultural representations of elephants together with a word "elephant" in one's native language, this symbol comes to refer indexically to those things, and the word "elephant" comes to be the name of that symbol, through which we can frame thoughts and statements about elephants. Any attempt therefore to define the meaning of the word will be doomed to failure, since the word just means ELEPHANT, and it derives the non-trivial part of its meaning through its reference relation to that (self-perpetuating) class of things in the real world.

This is not the place to explore the different accounts of how the appropriate indexical relation is established, or how the atomic concept gets associated with the prototype (or stereotype as it is sometimes called) of a concept. The key advantage of atomism is that it makes a clean distinction between the concept and what we know

about it. A descriptivist account of the contents of a concept will involve some set of features (i.e. broadly speaking a prototype). As a result if you and I differ in whether we think one of the features applies, then in effect we have *different concepts*. Given the data on instability described earlier, it would then follow that like stepping in the proverbial river, we never access the same concept twice. An answer needs to be given to this challenge, but (happily) space and time do not permit such a venture at this point. Just to note that it is a challenge not just to prototype theory but to all of cognitive science in as much as the latter aims to individuate concepts by representing conceptual contents (Fodor, 1998).

VII. Conclusions

I have covered a considerable amount of ground in this discussion of current issues concerning prototype representations. I hope to have convinced the reader that in spite of the unpopularity in certain quarters of prototype theory as a serious contender for representing concepts, the phenomena of prototypes are still with us, and still in need of explanation. Four major types of phenomenon have been reviewed, all of which seem to fit best with the prototype theory:

a. Membership in conceptual categories is vague, not only because people don't know enough about the domain, but also because word meanings are flexible and cannot be pinned down.
b. Degrees of typicality within a category influence a wide range of cognitive processes – from category-based induction, through memory interference and sentence processing, to the treatment of aphasics, and variation in typicality is not just about familiarity or availability of concepts in memory, but about similarity to the rest of the class.
c. The problem of how to treat the semantics of generic sentences is one of major importance, and prototype theory is the only account of concept representation that explains why so much of our semantic knowledge takes the form of statements that are "typically" true, rather than having a universally quantified truth. It is also the only approach that explains the non-logical combination of concepts under different forms of linguistic connective.
d. Finally the difficulty of defining word meanings remains a live issue. Prototype concepts cannot by their nature be simply defined. The problem can be stipulated away by taking an atomistic view, or it can be pushed back a level by taking a deference view, but neither of these will ever be a complete account of how the individual brain is able to use its internal representation of concepts for understanding, thinking and talking about its world.

At various points, I have tried to bring into the discussion notions that are common currency in philosophy, such as the Externalist view of conceptual contents, and ideas of deference and conceptual atomism. The integration of philosophy, lexical semantics and psychology into a true cognitive science of concepts is still a rather distant goal. Not only are the methods of enquiry of the three fields very different, but the value placed on different kinds of evidence varies widely as do the intuitive assumptions that drive the development of theory. However the goal remains a crucially important one. It should be possible for example for philosophy to set interesting research agendas for psychology, and for the data from psychology and linguistics to pose theoretical challenges for philosophy. The final unifying theory of concepts will need to explain how people's use of language is vague, variable, generic and opaque, as well as explaining how concepts can be reduced to atomic symbols for the understanding of logical reasoning. After all, Gödel and Frege, Wittgenstein and Russell developed their notions of the logical forms of natural reasoning with just the same biological apparatus as the rest of us. The mistake is to take our ability to appreciate the logical necessity of simple arguments such as A^B →A as the paradigm case of thought that requires explaining. Our minds have evolved to find it much less effortful to run down the vaguely drawn channels characterized by the range of phenomena reviewed here. The central notion of a prototype remains at the heart of our understanding of this way of thinking.

References

Aarts,B., Denison, D., Keizer, E. & Popova, G. (2004). Fuzzy Grammar: A Reader. Oxford: Oxford University Press.

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

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442-481.

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 factors in categorization* (pp. 101-140). Cambridge: Cambridge University Press.

Berlin, B. & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

Bouchard, D. (1995). Fuzziness and categorization, section 1.5.1. of *The Semantics of Syntax: A Minimalist Approach to Grammar.* Chicago: University of Chicago Press. Reprinted in Aarts et al. (2004) op.cit. pp. 479-485.

Braisby, N.R. (2005). Flexibility and Pragmatics in Categorizing Natural Kinds. Paper presented at the 9th European Congress of Psychology. Granada, Spain, July.

Braisby, N.R. (1993). Stable concepts and context-sensitive classification. *Irish Journal of Psychology, 14*, 426-441.

Carlson, G.N. & Pelletier, F.J. (1995). *The Generic Book*. Chicago: University of Chicago Press.

Cohen, B. & Murphy, G. L. (1984). Models of concepts. *Cognitive Science, 8*, 27-58.

Connolly, A.C., Fodor, J.A., Gleitman, L.R., & Gleitman, H. (2003). *Why stereotypes don't even make good defaults.* University of Pennsylvania: unpublished MS. October.

Estes, Z. (2004). Confidence and gradedness in semantic categorization: Definitely somewhat artifactual, maybe absolutely natural. *Psychonomic Bulletin & Review, 11*, 1041-1047..

Fodor, J.A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Oxford University Press.

Fodor, J.A. (2000). Multiple Review of Concepts: Where Cognitive Science Went Wrong, Replies to Critics. *Mind & Language, 15,* 350-374.

Fodor, J.A. & Lepore, E. (1994). The red herring and the pet fish: why concepts still can't be prototypes. *Cognition, 58*, 253-270.

Frege, G. (1903/1970) Concepts. In P.Geach and M.Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege,* Oxford: Blackwell, 1970. Reprinted in Aarts et al. (2004) op.cit., p. 33.

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought.* Cambridge MA: MIT Press.

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. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15,* 55-71.

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

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

Hampton, J.A. (1997a). Associative and Similarity-based Processes in Categorization Decisions. *Memory & Cognition, 25,* 625-640.

Hampton, J.A. (1997b). Conceptual combination. In K.Lamberts & D. R. Shanks (Eds.), *Knowledge, Concepts and Categories* (pp. 135-162). Hove: Psychology Press.

Hampton, J.A. (1997c). Psychological representation of concepts. In M.A.Conway (Ed.), *Cognitive Models of Memory* (Hove: Psychology Press.

Hampton, J.A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition, 65*, 137-165.

Hampton, J.A. (2004). *Reasons for Vagueness*. Paper presented at the Annual Meeting of the Psychonomic Society, Kansas City, November.

Hampton, J.A. (2005). Typicality, Graded Membership and Vagueness. City University, London: unpublished MS.

Hampton, J.A., Dubois, D. & Yeh, W. (in press). Effects of Classification Context on Categorization in Natural Categories. *Memory & Cognition.*

Hampton, J.A. & Estes, Z. (2000). *Appearance versus Reality: How Essential are Essences?* Paper presented at the Annual Meeting of the Psychonomic Society, New Orleans, November.

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

Hampton, J.A. & Jönsson, M. (2005). *Effects of noun modification on the plausibility of attribute information.* Paper presented at 46th Annual Meeting of the Psychonomic Society, Toronto, November.

Henle, M. & Michael, M. (1956). The Influence of Attitudes on Syllogistic Reasoning. *The Journal of Social Psychology, 44,* 115-127.

Homa, D. (1984). On the nature of categories. In G.H.Bower (Ed.), *The psychology of learning and motivation, vol. 18* (pp. 49-94). New York: Academic Press.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 418-439.

Kalish, C. W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory & Cognition, 23*, 335-353.

Kamp. H. & Partee, B. (1995). Prototype theory and compositionality. *Cognition, 57,* 129- 191.

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

Keller, D. & Kellas, G. (1978). Typicality as a dimension of encoding. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 78-85.

Kempton, W. (1978). Category grading and taxonomic relations: A mug is a sort of cup. *American Ethnologist, 5,* 44-65.

Kiran, S. & Thompson, C. K. (2003). The Role of Semantic Complexity in Treatment of Naming Deficits: Training Semantic Categories in Fluent Aphasia by Controlling Exemplar Typicality. *Journal of Speech, Language, and Hearing Research, 46,* 608-622.

Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Chierchia, G., & Link, G. (1995). Genericity: An Introduction. In G.N.Carlson & F. J. Pelletier (Eds.), *The Generic Book* (pp. 1-124). Chicago: University of Chicago Press.

Lakoff, G. (1987) *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.
Larochelle, S., Richard, S., & Soulières, I. (2000). What some effects might not be: The time to verify membership in "well-defined" categories. *Quarterly Journal of Experimental Psychology, 53A (4),* 929-961.
McCloskey, M. & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory & Cognition, 6*, 462-472.
McCloskey, M. & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology, 11,* 1-37.
McNamara, T. P. & Sternberg, R. J. (1983). Mental models of word meaning. *Journal of Verbal Learning and Verbal Behavior, 22*, 449-474.
Medin, D., & Atran, S. (2004). The Native Mind: Biological Categorization and Reasoning in Development and Across Cultures. *Psychological Review, 111*, 960-983.
Medin, D. L. & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.
Mellor, D. H. & Oliver, A. (1997). Introduction. In D.H.Mellor & A. Oliver (Eds.), Properties (pp. 1-33). Oxford: Oxford University Press.
Millikan R. (1998). A common structure for concepts of individuals, stuffs and real kinds: More mama, more milk and more mouse. *Behavioral and Brain Sciences, 21,* 55-66.
Murphy, G. L. & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.
Murphy, G. L. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.
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, 9*, 35-58.
Osherson, D.N. & Smith, E.E. (1982). Gradedness and conceptual conjunction. *Cognition, 12*, 299-318.
Osherson, D.N. & Smith, E.E. (1997). On typicality and vagueness. *Cognition, 64,* 189-206.
Posner, M. I. & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353-363.
Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge MA: MIT Press.
Putnam, H. (1975). The meaning of 'meaning'. In H.Putnam (Ed.), *Mind, Language, and Reality: Philosophical Papers, Volume 2, pp 215-271.*. Cambridge: Cambridge University Press.
Quine, W. V. (1948). On what there is. *Review of Metaphysics, 2,* 21-38. Reprinted in W.V.Quine, *From a Logical Point of View.* Cambridge MA: Harvard University Press, 1953, pp 1-19.
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.
Rips, L. J. (1989). Similarity, typicality, and categorization. In S.Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). Cambridge: Cambridge University Press.
Rips, L. J. (2001). Necessity and Natural Categories. *Psychological Bulletin, 127,* 827-852.
Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12*, 1-20.
Rosch, E.R. (1978). Principles of categorization. In E.R.Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27-48). Hillsdale NJ: Erlbaum.
Rosch, E.R. & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.
Rosch, E. R., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 491-502.
Ross, J.R. (1973). Nouniness. In Osamu Fujimura (Ed.), *Three dimensions of Linguistic Research.* Tokyo: TEC Company Ltd. Reprinted in Aarts et al. (2004) op.cit. pp. 351-422.
Roth, E.M. & Mervis, C.A. (1983). Fuzzy Set Theory and Class Inclusion Relations in Semantic Categories. *Journal of Verbal Learning and Verbal Behavior, 22,* 509-525.
Russell, B. (1923) Vagueness. *Australasian Journal of Philosophy and Psychology, 1,* 84-92. Reprinted in Aarts et al. (2004) op.cit. pp. 35-40.
Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22*, 189-228.
Smith, E.E. & Sloman, S.A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22,* 377-386.
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, J.D. & Minda, J.P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 3-27.
Sutcliffe, J.P. (1993). Concept, class, and category in the tradition of Aristotle. In I.van Mechelen, J.A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and Concepts: Theoretical Views and Inductive Data Analysis* (pp. 35-65). London: Academic Press.
Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293-315.
Verbeemen, T., Storms, G., & Verguts, T. (2004). Similarity and taxonomy in categorization. In K. Forbus, D. Gentner, & T. Regier, (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pp. 1393-1398. Mahwah, NJ: Erlbaum.
Wierzbicka, A. (1972). *Semantic Primitives*. Frankfurt: Athenaum-Verlag.
Wierzbicka, A. (1984). Apples are not a kind of fruit: The semantics of human categorization. *American Ethnologist, 11*, 313-328.
Wierzbicka, A. (1985). *Lexicography and Conceptual Analysis*. Ann Arbor, Michigan: Karoma.
Wierzbicka, A. (1987). English *Speech Act Verbs: A Semantic Dictionary*. New York: Academic Press.

Wierzbicka, A. (1990). "Prototypes save": on the uses and abuses of the notion of "prototype" in linguistics and related fields. In Tsohatzidis.S.L. (Ed.), *Meanings and Prototypes: Studies in Linguistic Categorization* (pp. 347-367). London: Routledge.
Williamson, T. (1994). *Vagueness.* London: Routledge.
Winman, A., Wneerholm, P., Juslin, P., & Shanks, D. R. (2005). Evidence for Rule-Based Processes in the Inverse Base-Rate Effect. *Quarterly Journal of Experimental Psychology, 58A*, 789-815.
Wittgenstein, L. (1953). *Philosophical Investigations*. New York: Macmillan.
Zadeh, L. (1965). Fuzzy sets. *Information and control, 8*, 338-353.