Cognitive Science 31 (2007) 355–383

Typicality, Graded Membership, and Vagueness

James A. Hampton

Psychology Department, City University, London

Received 25 May 2004; received in revised form 12 September 2006; accepted 20 September 2006

1. Vagueness in logic and psychology

A vague concept is a concept that picks out a category with no clearly defined boundary (see Keefe & Smith, 1997 for a review of theories of vagueness). Classical examples are the concepts labelled by adjectives such as BALD, TALL, or RED, and nouns such as CHAIR or VEGETABLE (Rosch, 1975). There is arguably no precise height at which a man or woman becomes tall, and so the class TALL MEN is not a well-defined set; consequently the truth of a statement such as "John is tall" may be in some sense undecidable if John is of intermediate height. Vague concepts are also susceptible to Zeno's infamous sorites paradox. If a man who is 1 m tall is clearly not tall, then nor is a man who is 1.0001 m tall. In general, if a man who is x meters tall is clearly not tall, then nor is a man who is $(x + 0.0001)$ m tall. But repeated application of this deduction can be used to prove that there are no tall men. Equivalently the argument can be reversed, starting with a man who is 2 m tall and working down the scale to prove on the contrary that all men are tall. Vagueness poses a serious challenge to theories of the logic of conceptual thought, and thus to theories of cognitive science—particularly those in the symbol-processing, representational theory of mind tradition.

The problem of vagueness has long been acknowledged as posing serious difficulties for philosophy and epistemology. How, it is asked, can we claim to have certain knowledge of the world when the very words that we use to express that knowledge are prone to such vagueness? Not only does vagueness cause difficulties with theories of reference—how it is that our concepts refer to classes of entities in the world—but it also creates major problems for the development of cognitively plausible logics that are consistent with conceptual thought. Logics that incorporate vagueness have been proposed (Zadeh's, 1965, Fuzzy Logic being the most well-known), but with limited success, and they have proved of little value as accounts of the psychology of vague reasoning (see for example, Osherson & Smith, 1981; Cohen & Murphy, 1984; Hampton, 1988b).

Vagueness has become an important issue in cognitive science with the development of psychological models of concepts and word meanings that take vagueness to be a central

Correspondence should be addressed to James A. Hampton, Psychology Department, City University, Northampton Square, London EC1V OHB. E-mail: J.A.Hampton@city.ac.uk

characteristic of human thinking. In particular, vagueness is closely associated with the psychological theory of conceptual representation known as prototype theory. This theory (Rosch, 1975; Hampton, 1979, 1995, 2006) holds that our concepts are structured on the basis of their similarity to some central prototype representation. The theory accounts for a whole range of "prototype effects." Items close to the prototype (typical items) are processed more rapidly, learned more quickly, remembered better, and so forth, than are atypical items (see Murphy, 2002). It also accounts for how people may consistently make categorization and reasoning judgments that are at odds with logic (Sloman, 1998; Hampton, 1982, 1988a,b, 1996, 1997a; Jönsson & Hampton, 2006;). Vagueness in psychology has, therefore, been embraced as an important source of evidence of the particular way in which the mind represents the world (for its importance in cognitive linguistics, see Lakoff, 1987).

These two different views of vagueness—on the one hand as a philosophical embarrassment, and on the other hand as a central plank of psychological and cognitive linguistic theorising—have naturally led to controversy between the different branches of cognitive science. The controversy is seen at its clearest in the rejection of the notion of concepts as prototypes by Fodor and Lepore (1996), (see also Fodor, 1998). According to Fodor and Lepore, taking concepts to be prototypes precludes the possibility of a successful account of the compositionality of concepts in logical expressions, given that prototypes do not show compositionality of the right kind. And compositionality for them is a nonnegotiable assumption of the representational theory of mind which lies at the heart of cognitive science.

In this paper I seek to explicate the role of prototype theory in providing an account of both vagueness and the problems of logical reasoning with vague concepts. I begin by considering two influential proposals that challenge the role of prototypes in determining vagueness. These two important papers—by Kamp and Partee (1995) and Osherson and Smith (1997), suggest that a distinction should be made between the psychological phenomena of typicality on the one hand, and the logical phenomenon of graded membership or vagueness on the other hand. I argue that there are major difficulties in sustaining such a distinction. In the following sections I present a psychological account of vagueness based on prototypes, and suggest ways in which it could be made to handle the problems of the logic of vague predicates. I do not address other known objections to prototype theory, such as the influence of essentialist beliefs on categorization (Rips, 1989), or dissociations between typicality and categorization (Rips, 1989; Ahn & Dennis, 2001; Hampton, 2001; Thibaut et al., 2002). The different ways in which conceptual information is represented, and the different ways in which it is used for diverse purposes such as rating typicality, drawing inductive inferences, naming or categorizing the world require a much more detailed account than the model given here. The model simply serves to illustrate how typicality effects and issues of vagueness of category membership both arise from the particular characteristics of our conceptual representation system. I also do not attempt to solve the metaphysical problems of vagueness (for example whether it is the world itself, or concepts, or word meanings, or language use, or all of the above where the problem lies). My aim is to show that psychological intuitions of typicality, category membership, truth and vagueness are explainable in terms of a theory of how concepts are represented and combined, and that the logical problems with psychological models

identified by Kamp and Partee (1995) and Osherson and Smith (1997) can be answered in these terms.

The final section of the paper considers how individual intuitions of vagueness may arise from meta-beliefs about language use in one's community.

2. Distinguishing typicality and vagueness

The importance of vagueness in the problematic relation between the philosophical and the psychological approach to concepts was recently addressed in two papers in *Cognition* by Kamp and Partee, 1995, (K&P), and by Osherson and Smith, 1997, (O&S). A central argument in both papers rested on a proposed distinction between gradedness of membership c^e and goodness of example c^p. (The notation, based on logical characteristic functions c, is taken from K&P, and was also adopted by O&S. Since it is not particularly transparent to the general reader, I shall henceforth refer to gradedness of membership, c^e, more simply as M for membership and goodness of example, c^p, as T for typicality. They are intended to be coreferential with K&P's terms.)

Vagueness generally relates to the question of whether or not, and to what degree an instance falls within a conceptual category. Whereas some concepts such as PRIME NUMBER correspond to categories with clear-cut boundaries in a particular domain, vague concepts such as BALD or RED allow for graded category membership, where at the boundary region it may be neither clearly true nor clearly false to say that an instance is in the category. Accordingly gradedness of membership, or more precisely the degree of membership, M, is defined as an index of this gradation, with continuous values ranging from 0 for clear non-members through to 1 for clear members. (Well-defined concepts would be a subset of concepts for which M takes only the values of 0 or 1.)

Typicality, T, on the other hand, reflects how representative an exemplar is of a category. Robins are to most English speakers more typical of birds than are penguins (Hampton & Gardiner, 1983; Rosch, 1975), and this fact is attributed (in large part) to robins sharing more features in common with a prototype representation of the category birds than do penguins. The measure T is therefore primarily an index of the degree of similarity of an instance to the category prototype. (The same index T could also be based on similarity to stored exemplars, as in Medin and Schaffer's (1978) context model. For more details of the different models see Hampton, 1997b; Murphy, 2002, pp. 41–65).

Establishing the proper basis for the distinction between these two measures is of particular importance, since if, as O&S and K&P contend, typicality T has quite different properties as a variable from degree of membership M, then similarity to prototypes will not be able to account for both measures. If they are right, typicality T may turn out to be a dimension of purely psychological interest, responsible for the range of typicality effects but of little value for explaining the role of concepts in determining the truth of sentences, while degree of membership M is more a matter of concern for logicians and ontologists. In this and the following section the arguments presented by O&S and K&P are examined in more detail.

2.1. Osherson and Smith (1997)

O&S present three major arguments for the distinction between degree of membership and typicality. All three rely on showing that differences may be observed in T in the absence of corresponding differences in M.

O&S first point to the commonly made observation that two objects may both be clear members of a category (a robin and a penguin are both indisputably birds) and yet may also clearly differ in their typicality. Such examples led O&S to reject the central tenet of prototype theory—namely that the same underlying dimension of degree of resemblance that gives rise to intuitions of typicality, T, is also involved in determining category membership, and so gives rise to differences in degrees of membership in the category, M.

Put simply, we can tell by the almost universally positive categorizations of both robins and penguins as birds, that their graded membership M as birds must be 1,

$$M_{bird}(robin) = M_{bird}(penguin) = 1 \tag{1}$$

whereas, in terms of typicality T, since robins are more typical than penguins as birds,

$$T_{bird}(robin) > T_{bird}(penguin) \tag{2}$$

It follows that M and T are different functions. From this O&S conclude that membership and typicality reflect different underlying psychological processes.

In addition, O&S present two arguments that whereas graded category membership may be sensibly mapped onto the closed interval from 0 to 1, typicality is better thought of as unbounded. The first has to do with the degree of match between an instance x and a category C. In particular O&S offer the following principle.

> Suppose that object O illustrates a category N named by a noun, and also a category A named by a potential modifier of N. Then O is often a better example of the modified category AN than it is of N. (3)

For example, if one were to find an apple that was perfect in every respect at matching the prototype for apples, and it also happened to be a perfect red, then it would be an even better match for the prototype of RED APPLE than it was for APPLE alone (see Smith, Osherson, Rips & Keane, 1988, for evidence of this effect, and Storms, de Boeck, van Mechelen & Ruts, 1996, for an extension to noun class conjunctions). Since principle (3) can be applied recursively (by finding another modifier adjective A′ to generate the conjunctive concept A′ ∩ AN), it follows that unless there is some constraint on the number of possible modifiers, there can be no maximum constraint on typicality. Thus the perfect red apple, should it also be very round and very shiny would be even more typical of the concept ROUND SHINY RED APPLE. It therefore cannot have already been as typical as it is possible to get for RED APPLE. Thus typicality T has no upper bound.

The parallel argument presented by O&S applies to concepts with an inherent dimensional component, such as HOT, GIANT or BULLY. Such concepts involve features that Barsalou (1985) identified as "ideals." Instead of typicality being related simply to the central tendency of the category, as happens for many concepts, with these dimensional concepts typicality will continue to increase as an instance becomes more and more extreme on the idealised

dimension. All other things being equal, the more aggressive and hateful an individual becomes, the more typical they are of the category bully. Such cases therefore represent a second situation in which typicality T may be unlimited at the upper bound.

Having argued that T may (at least in these two cases) be unbounded, O&S then point out that since degree of membership M lies in the range from 0 to 1, whereas T has no maximum value, once again typicality cannot be the same as degree of membership.

2.2. The threshold model

O&S's first argument is invalid for the following reason. Two measures may relate to the same underlying dimension in a perfectly regular way, without having the same range or discriminatory power. Functions (4) and (5) are both simple functions of the same variable x, but whereas the value of function (4) for any real number x is bounded within the open interval (0,1) between 0 and 1, (5) is bounded at 0 but unbounded at the top end of the scale as x increases.

$$f(x) = e^x/(1 + e^x) \tag{4}$$

$$g(x) = e^x \tag{5}$$

Yet they are simply related to each other, and both entirely determined by the same variable x. Pointing to differences in the mathematical properties of T and M is not therefore a valid argument for their reflecting very different underlying psychological processes. As similarity to a prototype increases, there is no reason why M should not initially remain at zero, then rise in a monotonic fashion to one, and then remain at one as similarity approaches its maximum. At the same time, T could be more sensitive to changes in similarity both for clear non-members (where M is zero) and clear members (where M is one).

I propose (contrary to O&S's conclusions) that both degree of membership M and typicality T are based on a single underlying metric of similarity. It could be similarity to stored exemplars (Medin & Shaffer, 1978; Nosofsky, 1988), similarity to an abstract representation of central tendency (Hampton, 1995; 1998), or similarity invol causal principles and relational properties (Murphy & Medin, 1985; Ahn & Dennis, 2001)—it makes little difference to the argument. (For certain dimensional concepts, however, similarity has to be supplemented by closeness to an ideal—see discussion later in this section.) Then what we observe or measure as ratings and categorization judgments are two types of behavior, both based on the same single underlying notion.

Intuitions of typicality T are a direct monotone function of underlying similarity, and one which (for the sake of argument) may not reach an asymptote at the top end of the scale. If the similarity match of an instance to the prototype is increased (for example by adding additional specifications to the prototype as O&S proposed), then so is the typicality T. Judgments of degree of membership M on the other hand follow a different function, one which asymptotes at 1 or 0 at the top and the bottom of the scale respectively. Hence there is no contradiction between expressions (1) and (2). In other words, just because $M \neq T$, it does not follow that M and T may not both be nondecreasing functions of a single underlying dimension of similarity, and hence related by some function (for example from T to M).

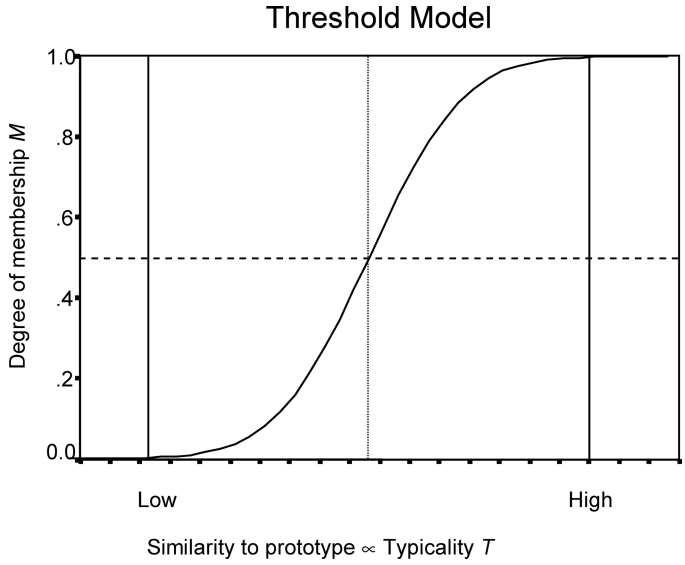

Fig. 1. Illustration of the relationship between typicality T and graded membership M.

A model based directly on prototype theory (Hampton, 1979, 1995; Rosch, 1975), the threshold model, links the two measures directly. The model uses two different nondecreasing functions of the same similarity measure:

$$T = t(sim)$$
$$M = m(sim) \tag{6}$$

Whereas the function t may be unbounded, the function m for this model lies in the open region (0,1). (I will argue later that it probably makes better sense for M to lie in the closed region [0,1] which includes values of 0 and 1.)

Figure 1 shows a putative relation between similarity to prototype and graded membership. For simplicity, a linear relation has been assumed between similarity to prototype and typicality T. The function for M shown here is the cumulative normal distribution. The region within the solid vertical lines is the borderline region in which membership is noticeably graded. However there are no discontinuities in the function, so we can neatly avoid the problem of second order vagueness associated with dividing the scale into clear and borderline regions. The function is continuously graded, and simply asymptotes at 1 or 0. (The positioning of the vertical lines carries no theoretical weight here, but serves merely an illustrative function).

To show that this is not simply a theoretical abstraction, Fig. 2 shows results from an analysis of McCloskey and Glucksberg's (1978) data reported in Hampton (1998). The data reflect independent judgments of typicality and binary categorization decisions for 482 items in 17 categories. In this figure, degree of membership M, shown on the vertical axis, is indexed by the probability of a positive categorization,[1] while mean rating of typicality is shown on the horizontal axis. Error bars show 95% confidence intervals. The theoretical model that assumes that M follows a cumulative Gaussian function of similarity and T a

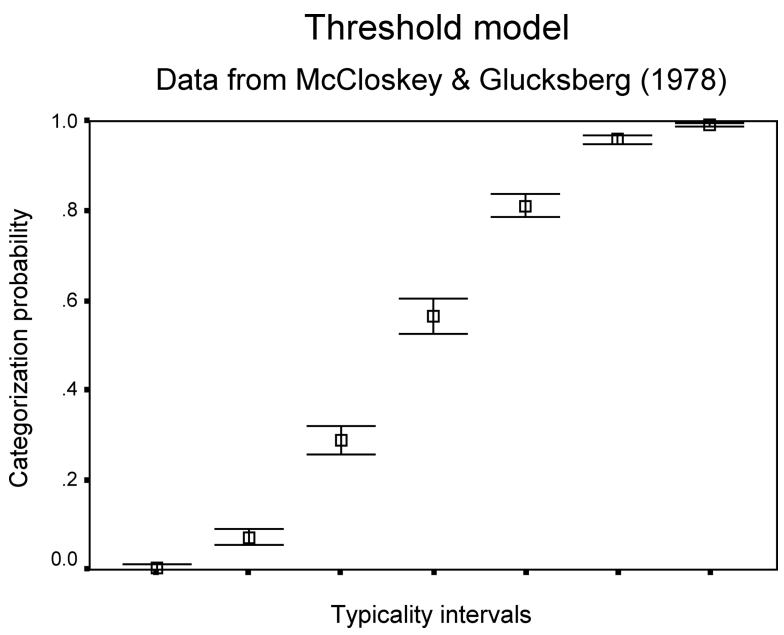

Fig. 2. Illustration of the relation between ratings of typicality and yes/no judgments of category membership.

linear function of similarity clearly provides an excellent first approximation to the observed data.

The threshold model therefore is easily able to account for the case of robins and penguins. Both have sufficient similarity to the concept of bird (however that is represented) to pass the threshold for category membership. But robins have higher similarity than penguins and so achieve a higher score for typicality. For the case of the ROUND SHINY RED APPLE, there is again no problem with M remaining at 1 as T increases. Note also that O&S's example involves comparing a given object with a sequence of increasingly specific concepts. It is therefore possible for there to be a maximum value of T for any given one of those *concepts*—even if there may be no maximum value of T for a given *object*. Whereas a concept (according to most descriptivist accounts in psychology) is a finite representation stored in our finite memories, an object is not a representation at all. An object can be described in an infinite number of ways—hence there is no theoretical limit to the specificity of a representation of that object in any given situation. The tentative conclusion then is that for any given concept there may be one or more objects that achieve maximum possible typicality, but for any given object there will be no concept for which that object achieves maximum typicality. (Empirical evidence would of course be useful to test this conclusion).

In order to counter O&S's final example, it is necessary to make a division between two types of concept. Their first demonstration (the RED ROUND SHINY APPLE) involves increasing levels of specification for a given prototype. Typicality in such concepts will therefore depend simply on similarity, or on how well the specific descriptors match the object. The closer an object fits the cluster of concepts in the nominal phrase, the more typical it will be. Their second demonstration involves dimensional concepts such as TALL or BULLY, where there is

perhaps no end to the degree to which an item could match the prototype (since the most typical example would have to reach an unattainable ideal). Concepts like BIG, TALL or VIOLENT differ from concepts like RED, ROUND or HAS FOUR LEGS in that (as is well known) categorical judgments are to a much greater degree dependent on establishing an appropriate context. A big ant is no match for a small elephant, a tall 5-year-old is smaller than a short 18-year-old, and a violent hamster may be a lot less violent than a friendly pit-bull terrier. Typicality and membership for such concepts, or for noun concepts like GIANT or BULLY that incorporate dimensional ideals, will inevitably involve determining not how close to the average of the class an object is, but how far along the scale the object is. Barsalou (1985) showed that even for common semantic categories like VEHICLE or WEAPON, proximity to an ideal (in this case efficiency or effectiveness in achieving its function) played a role in determining typicality. Where there is a dimensional ideal, typicality will increase along that dimension, but so will membership.[2]

Because they contain dimensions with no upper bound, some dimensionally based concepts may have no upper bound for typicality T. In the meantime, membership M will often be ill-defined without a more specific context being provided. The concept TALL will have no exemplar with maximum typicality, given that TALL buildings can go on increasing in height indefinitely. Membership in the category TALL however will reach values of zero and one as height varies relative to the relevant standard in any given context. Once again, these facts are quite consistent with both measures being based on the same underlying variable, although in this case the variable will reflect both similarity to the central tendency of the class and extremity on the dimension for which an ideal has been specified. The students in my class may have an infinite ability to exceed any given level of excellence I might choose—so that their typicality in the category GOOD STUDENT could be unbounded. Yet I have no difficulty in awarding pass/fail grades on exactly the same basis as I would award differential grades to reflect degrees of excellence. There is no logical contradiction in having one measure that is largely binary (although with the possibility of a vague boundary region) and another measure that is continuous, and for both to be based on the same underlying continuous variable.

Consider one last example—the classification of a particular note on the piano as "high" or "not high." At low frequencies in the bass range of the piano, the note is clearly not high. Hence M is zero. Around the middle third of the keyboard, the note becomes increasingly likely to be called "high," while above some point, it is clearly "high" to anyone who is asked and is familiar with the range of a piano, so $M = 1$. Furthermore, let us suppose that an experiment reveals that people judge the typicality of a note in the category HIGH NOTE in a way that rises strictly monotonically with the actual pitch of the note. That is HIGH involves an ideal in its prototype. Now two notes at the top end of the scale may be equally clearly in the "high" category. Under normal circumstances of hearing etc., no one could doubt the classification of both as high, so $M = 1$ in both cases. Yet everyone would also agree that one note is higher than the other, and would therefore judge that T is greater for one than for the other. Moreover, whereas degree of membership M in the category lies in the interval [0,1], the pitch of a note (and hence T) may continue to rise (at least to the limit of our hearing, and, in principle, indefinitely). (What would happen to typicality at the top of the scale is an empirical question, but presumably when the note is no longer audible it ceases to be a note at

all to the observer and is just a period of silence, so the question of its typicality or category membership no longer arises).

Both of the arguments used by O&S against prototypes determining graded membership apply equally well here, and if their arguments are valid we are led to the conclusion that the classification of a note as in the category "high" uses some information other than whatever determines its typicality as a high pitch. Since pitch is unbounded and can vary within clear members of the category of high notes, whereas degree of membership in the class "high note" is bounded and remains constant above some level, the classification of notes into high and not-high cannot be based on the same underlying dimension as typicality.

But this is clearly wrong. For one thing, the higher a note in the middle region, the more likely it is to be judged as "high," just as, within the borderline region of a concept, those instances with greater similarity to others in the class are more likely to be placed in the category. More importantly, we can actually anchor both psychological behaviours (classification and typicality of the pitch in relation to the vague term) to a measurable physical dimension—the fundamental frequency of the note. My claim is that such a dimension is also available in the conceptual realm—the degree of match between a mental representation of a class or an individual and the relevant features or dimensions contained within a concept representation.

In the discussion thus far, I have argued that M is at zero or one on the basis of whether people would all agree that the item lies in the category or not. This operationalization of M has the clear advantage of being easily measurable, and also has intuitive appeal as a way of cashing out the quasi-philosophical notion of rationality as in "Any reasonable person would have to agree that . . . ." If any individual refused to accept that a robin was a bird, we would have no qualms about concluding that they lacked a correct grasp of one or other concept, or that something was seriously interfering with the communication between us (maybe they were being blackmailed, or were trying to ruin our experiment for other reasons).

How reasonable is it to take probability of categorization as a measure of M? As with any empirical measure, some agreed standardized procedure for obtaining category judgments under suitably controlled conditions would first need to be put into place, but let us assume for the moment that this would be unproblematic. Probability of categorization certainly has the right distributional properties. It reaches 1 at a point below the maximum for the typicality scale, so that (for example) both robin and penguin would have values of 1. Yet how should one interpret the fact that if asked if chess is a sport, 60% of a sample say yes, and 40% say no? Can this be directly mapped to an M of 0.6? Although tempting in its simplicity, one should guard against this mapping. It may be that each individual is uncertain how to classify chess as a sport and responds probabilistically, with a 0.6 probability of saying yes. Alternatively however each person may have a clear idea of whether chess is a sport or not, and it so happens that the firm belief that chess is a sport is held by 60% of the sample. More likely still is some mixture of both these accounts. McCloskey and Glucksberg (1978) provided evidence on this question by retesting subjects' categorization decisions after a period of a few weeks. They found that subjects were likely to change their categorization decisions on the second occasion, and did so more often for the items of borderline typicality. Levels of inconsistency were however lower than levels of disagreement between one person and another, showing that there was a certain degree of consistent individual difference in how the items were categorized. It is not therefore the case that the categorization probability just

reflects the distribution of firmly held beliefs in the population, nor is it the case that it just reflects the likelihood that any one individual will give a positive categorization. I return again to the question of how best to interpret the theoretical meaning of M in the final section.

An unresolved question relates to what I believe is the fundamental intuition driving O&S's first argument—namely that a penguin and a robin just both are fully and indisputably birds. The issue comes down to whether M should be thought of as falling within the closed interval [0,1] which includes actual values of 0 and 1, or whether [as in the mathematical example given in expressions (4) and (5)] it falls within (0,1) and so never actually reaches complete membership or complete nonmembership. Naïve intuition would clearly have great difficulties with the open option. Not only do we have a sense that there can be no doubt (however slight) about a robin being a bird, we would find it even harder to accept that M would never reach zero if there were the slightest identifiable similarity between an item and a concept. Suppose that part of the prototype for FURNITURE is the property of being a physical object. Then one would predict that all physical objects would be likely judged more typical of furniture than all non-physical objects—mountains are more typical furniture than the smell of roses. Perhaps people might agree to this, but they would surely object to the corollary that in some sense mountains are therefore in the category of furniture to some infinitesimal degree. If M takes values only within (0,1), then by the threshold model, (7) will hold:

$$\text{For any items } (x, y) \text{ and any concept } A \text{ such that } T_A(x) > T_A(y), \text{then } M_A(x) > M_A(y) \tag{7}$$

As a consequence any difference in typicality entails a difference in membership. There is a strong case for saying that (7) violates our intuitions both for cases of clear membership and for cases of clear non-membership.

Unless one is happy to ride roughshod over these intuitions (with the risk that the notion of M loses much of its intuitive appeal), the value of M will have to lie within the closed range [0,1]. Mountains are furniture to degree zero, and robins are birds to degree 1. The question as it relates to vagueness then becomes one of specifying precisely how well an object or class has to match a concept for M to attain a value of 1 (and how badly it has to match for M to be 0). Note first that there is no mathematical reason why the hypothetical function for M in Fig. 1 should not reach 0 and 1 "smoothly" at the limits of the central borderline region. For example as the value of x in the simple quadratic function x^2 approaches zero from above, so the slope of the function also tends to zero. The function in (8) (see Fig. 3) would therefore serve the purpose of a smooth[3] function that lies within [0,1] for all values of similarity S:

For Similarity S, we define: S_H = the upper bound to the borderline region above which $M = 1$, S_T = the threshold where M is 0.5, and S_L = the lower bound below which $M = 0$. In addition let S_T lie midway between S_H and S_L. Then the function M is defined as:

where

$$S_L > S, : M = 0$$

$$S_T > S > S_L : M = 2\,[(S - S_L) / (S_H - S_L)]^2$$

$$S_H > S > S_T : M = 1 - 2\,[(S_H - S) / (S_H - S_L)]^2$$

$$S > S_H : M = 1 \tag{8}$$

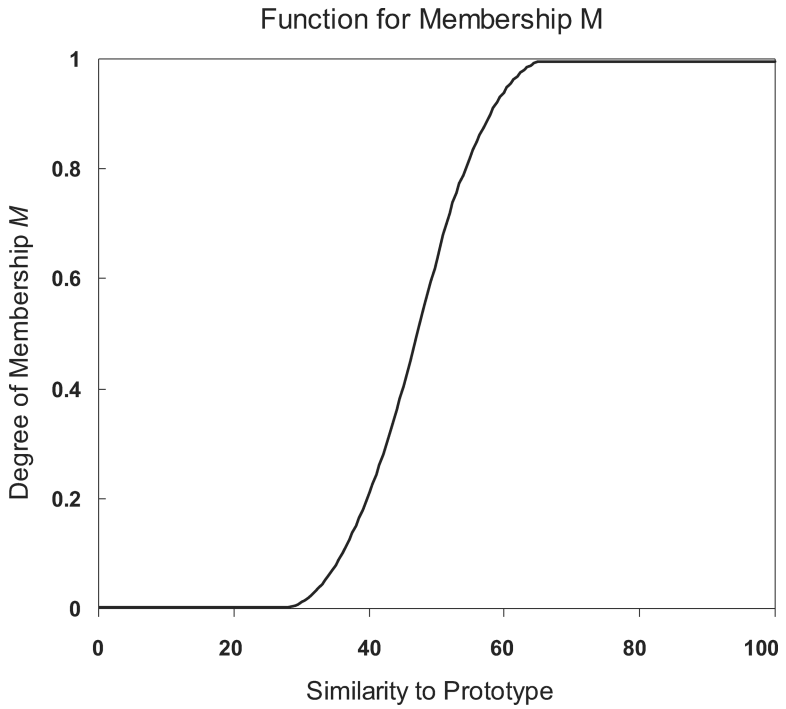

Fig. 3. Smooth threshold function for membership M based on a hypothetical similarity scale, using the function defined in (8) in the text, where M actually reaches 1 and 0 rather than asymptoting to those values.

In practice of course no precise values can be given for the higher and lower boundaries of the vague region on the similarity scale, nor is it required for the threshold model that such values be precise. In principle, within a sufficiently well-constrained context (for example just one individual tested repeatedly with very clearly specified materials), the prototype and similarity scale used by that individual could be discovered, and the limits of the borderline region within which inconsistent categorization occurs for that individual could be delineated. In practice the determination of the upper and lower limits of the vague region will be as vague as the determination of the 50% threshold point. Both depend on the distribution of threshold values across categorizations and the inherent variance in the scale values themselves, and so would resist a precise determination.

O&S end their paper with a challenge, asking what it might mean to be able to justify a value of M of 0.85 as opposed to 0.91 for some categorization. The interpretation of M is discussed further in section 4, but we have already seen that it should be closely related to the likelihood of a categorization for the following reason. If values of 1 and 0 for M are given to items for which no reasonable person could doubt that the item is or isn't a category member, then there should be a close mapping to probability of categorization at least at each end of the measure.

As yet we have no theory of measurement for gradedness or typicality that can yield interval level scaling such as would warrant the mapping of M onto real numbers. But this is not to

say that when more is known about how judgments of category membership are determined we would not be able to create one or more sensible measurement scales for M. The problem is the same as that faced by psychometricians aiming to capture some systematic difference between individuals. In the days before psychometric scales were developed, the very same question could be asked about, say, intelligence or depression. What would be the meaning of saying that a person is intelligent to degree 110, or depressed to degree 45? Without wishing to comment on the current state of psychometric science, once a principled way of measuring a construct has been discovered there is then no difficulty in replacing ordinal with interval scaling. To take a different example, a pre-scientific scholar in the era before the development of thermometers could have offered exactly the same appeal to intuition: "no one has yet explained to me what it might mean for an object's temperature to be 0.85 as opposed to 0.91; things may be hotter or cooler, and there may be ways to show this, such as whether wax melts or water boils, but it is nonsense to place an exact value on such an impression." Greater understanding of how the mind measures similarity and transduces that measure into categorization judgments should allow us to develop an appropriate scale for M, just as the invention of mercury or thermocouple thermometers allowed temperature scales to be defined.[4]

3. Logical combinations of prototypes: Kamp and Partee (1995)

Although concerned with a rather different set of issues, K&P also arrive at the conclusion that it is necessary to keep the bases of M and T distinct. Following Osherson and Smith's (1981) seminal challenge to prototype theory, K&P were particularly concerned to develop a logic for the combination of vaguely defined classes for which (unlike Zadeh's, 1965, fuzzy logic) the self-contradictory $A(x)\&\neg A(x)$—"x is an A and x is not an A"—and the tautological $A(x)OR\neg A(x)$—"x is an A or x is not an A"—should have truth values of 0 and 1, respectively, regardless of the fuzzy truth of the statement "x is an A."

Osherson and Smith (1981, 1982) demonstrated the problems of applying fuzzy logic to these propositions. In Zadeh's fuzzy logic, a conjunction has the minimum and a disjunction the maximum of the truth values of the two components. Suppose we anchor the midpoint 0.5 of the M scale to the point of equivocality between calling a proposition true and calling it false. Then if for some instance x and class A, $M(x, A)$ is 0.5, both $A(x)$and $\neg A(x)$ will have values of 0.5, and the two expressions above will both be evaluated as 0.5 true, an obviously counterintuitive conclusion to reach in each case.

K&P proposed to solve the problems of fuzzy logic's failure with self-contradictions and tautologies using supervaluation theory, (Fine, 1975; Kamp, 1975). In brief, this theory proposes that for any vague predicate, there is a partial model which assigns a value of true or false to clear cases at either end of the scale, and for the middle vague region leaves the truth undefined. For example for "X is tall," there would be a top region on the height scale, for which the proposition was always true, a bottom region for which it was always false, and a middle region, for which it was neither true nor false—its truth in this region would be undefined. The theory then seeks to evaluate the truth of a complex proposition by considering its truth within each possible completion of this partial model. A completion of the partial

model is achieved by assigning the values true or false to each instance in the middle region, in order to arrive at a particular model (or "precisification" in Fine's terms) in which it is clear for each instance whether it is in the class or not. The full set of such completions corresponds to all possible ways of assigning true or false to the set of borderline cases (although constraints may also apply to rule out some completions—see below). Now since $A(x)\&\neg A(x)$ has a value of 0 in every possible completion, and $A(x)OR\neg A(x)$ has a value of 1 in every possible completion, their truth value can be determined for all completions, as a supervaluation of truth values across all possible models. Effectively the method is saying "it doesn't matter whether you count this instance in the category or not; on those occasions where you do, $A(x)\&\neg A(x)$ must be false, and on those where you don't, $A(x)\&\neg A(x)$ will still be false, so it is false for all possible ways of categorizing the boundary cases."

K&P proposed an extension of the supervaluation approach in which a measure of degree of membership M (and hence something possibly equivalent to a fuzzy truth value for "x is an A") could be generated. They suggested that the degree of membership index M could be interpreted as the proportion of all possible completions of the partial model in which x is included in A. Now, if there are no constraints on completions, this value would always be 0.5 in the borderline region, since for every model in which a borderline item x is included in A there is exactly one equivalent model that is the same in every respect except that x is not included in A. But if we add partial ordering constraints amongst the elements in the domain, then some instances will be included in A in a higher proportion of completions than others. For example, membership of the category TALL in the domain of adult men might sensibly have an ordering constraint on individuals in terms of their height. Such a constraint would rule out any completion of the partial model in which an individual assigned to the category TALL had a (noticeably) lower height than an individual excluded from the category. As evidence that people do honor such constraints, Hampton, Estes, and Simmons (2005), reported that when participants made category judgments about each of a pair of colors on the blue-purple boundary the judgments obeyed the ordering constraint in the vast majority of cases (there were 8 violations in nearly 6000 trials).

Through this device one can then (in principle) map all possible values of height within the borderline range onto values of M between 0 and 1. The value of M for a sentence such as "A man who is 172 cm is tall" would be defined precisely as the proportion of possible heights, lying within some defined central borderline region (say 170–180 cm), whose value is less than 172 cm. If we determine M in this fashion, and calculate proportions as the ratio of distances on the height scale, then Fig. 4 illustrates the function that K&P's proposal would generate, mapping M onto the underlying scale. Of course this is not the shape of the curve shown in Fig. 2, but the shape of the latter curve may perhaps be the result of additional psychological processes applied to the supervaluation-based M.

K&P go on to note however, that the supervaluation account cannot play a role in explaining the intuitions of typicality T of instances or subclasses in a complex concept, since M maps only onto the closed interval [0,1], and does not permit of variation amongst category members that are all clearly in the category. Supervaluations can have nothing to say about T and so if supervaluations explain degree of membership M (and their treatment of $A(x)\&\neg A(x)$ and $A(x)OR\neg A(x)$ predicates suggests they do a good job there), then degree of membership and typicality will each need different accounts.

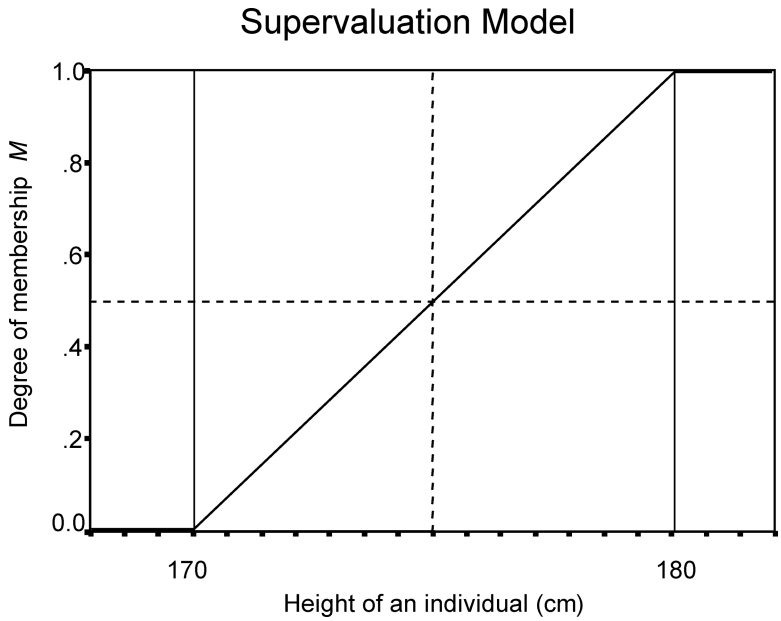

Fig. 4. Function relating degree of membership to typicality according to the supervaluation proposal.

The problems with K&P's arguments have to some extent already been discussed. In particular, restriction of vagueness to the central region of a scale begs the question of how the boundaries of the vague region are to be defined, if not in a vague manner. Presuming there is a clear region at the top and bottom of the scale means that we now need to specify two vague boundaries rather than just one. Of course the threshold model will also have to provide some account of where the vague region ends, but as discussed above it can do this in a vague way since (a) the function approaches 1 and 0 with zero slope,—there is no discontinuity,—and (b) it is assumed that the same factors that render the threshold criterion itself subject to variation apply equally to the top (S_H) and bottom (S_L) limits of the region. The function relating M to similarity is defined in the threshold model in Fig. 1 as the cumulative frequency distribution for placements of the threshold on the scale. The top limit (the limit of the vague region) is the maximum level that the threshold can reach within its distribution, which in practical terms will depend on all of the same individual and situational variables that affect threshold placement.

Even more problematic for K&P's distinction between M and T is the question of where the partial orderings within the boundary region would come from in order to derive M, in the case of concepts not based on single dimensions like height. For example why should tomatoes and avocados have a higher value of M as fruit than pumpkins or olives (as suggested by categorization probabilities reported by Hampton, Dubois, & Yeh, 2007)? To base the orderings on similarity relations would be to import information about the prototype that their model wishes to manage without. Yet short of a huge amount of arbitrary stipulation, there is no alternative way to account for consistent variations in degrees of membership. Indeed unlike colors, there is no monotonicity constraining the ordering of membership for non-dimensional

concepts like FRUIT. (In a recent unpublished study I found that simultaneous categorization of two borderline items for a semantic category were effectively independent of each other). So the supervaluation account cannot do the work of accounting for intuitions of vagueness unless it also incorporates a means of imposing constraints on the possible completions. But such constraints must be "soft" (i.e. probabilistic) and in the absence of any other possibility must be based on the same dimension of similarity as is assumed to underlie typicality.

3.1. Alternative accounts of A(x)&¬A(x) and A(x)OR¬A(x)

The threshold model proposes that *T* and *M* both depend on the same underlying dimension, even though *M* reaches a maximum well before *T* reaches its maximum value for any set of exemplars. So in order to maintain the view that degree of membership *M* depends on similarity to prototype, I need an alternative account to the supervaluation approach to degree of membership. In particular I need to solve the logical problems of the propositions *A(x)&¬A(x)* and *A(x)OR¬A(x)* that cause the difficulty for fuzzy logic, and that supervaluations so neatly avoid.

In the following section a number of different accounts are offered. These are not intended to be mutually exclusive, but rather to map out the territory of different ways in which the phrases may be understood and interpreted. To prepare the reader for the very different accounts offered, section 3.2 suggests that we can solve the problems because verbal phrases containing explicit logical connectives may be evaluated syntactically rather than semantically, and so fuzzy membership and prototypes are not brought into play. Section 3.3 suggests that the explicit conjunction or disjunction of complement sets might lead to a psychological supervaluationist treatment that provides the correct answers of true and false to the two problem cases because extensional reasoning is used. In contrast to these two approaches, the last approach in section 3.4 argues the case that in fact we may have alternative interpretations of the two phrases which do not take truth values of 1 and 0 respectively. Furthermore these intuitive interpretations can be explicated by looking at the results of experiments that have investigated how people combine prototypes with the logical connectives AND, OR and NOT. The extent to which each of these approaches turns out to be a correct description of people's behavior is an empirical question that will require further investigation.

3.2. Linguistic form makes the logical form transparent

My first account appeals to the difference between rule-based and similarity-based processing (e.g., Ashby et al., 1998; Sloman, 1996). When a verbal expression has a syntactic form that makes the logical form explicit, people are very ready to assent to it on the basis of the form alone. Thus *modus ponens* arguments such as (9) may be accepted as valid without any need to provide a meaning for the individual terms:

If all lokies pling then all maks flug.
All lokies do pling, so all maks must flug. (9)

It is possible then that the logically explicit forms of the *A(x)&¬A(x)* and *A(x)OR¬A(x)* assertions trigger an immediate recognition of the necessity of their truth and falsehood at

a purely syntactic level (just as the tautology "every snorl is a snorl" and the contradiction "there is a snork which is not a snork" can be evaluated without knowing what snorls or snorks are). Having no need to retrieve the prototype meanings of the terms, the issue of vagueness is thus by-passed. Effectively this would be the approach promoted by O&S when they argue that truth functions for concepts operate with Boolean logic.

This proposal is somewhat post hoc, since it is not clear how many such syntax-only driven inferences would occur to an average person. For example the apparently obvious set inclusion relation between the subject noun phrases in (10) is surprisingly difficult for people to notice and use:

All sofas have backrests
All uncomfortable handmade sofas have backrests. (10)

Jönsson and Hampton (2006) found that people showed a strong tendency to agree to the first sentence but disagree with the second. Even when the two were placed side by side and people asked if they were equally true or not, there was still a significant tendency to commit this fallacy, with many people choosing the more general statement as more true.

In general people find it very hard to separate the validity of the logical form of an argument from their beliefs in the truth of its statements (Evans et al., 1983). There are cases however where exposure to explicit logical form has been shown to reduce or eliminate nonlogical biases in reasoning (Sloman, 1998, Experiment 4). This account makes the testable prediction that judging the truth or falsity of sentences of the form $A(x)\&\neg A(x)$ and $A(x)OR\neg A(x)$ should be rapid and accurate, regardless of the semantic contents of A and x.

I have two further solutions to offer. One aims to solve the problem through a more psychologically plausible adaptation of supervaluation theory. The other solution considers the empirical data on how our conceptual system combines prototype concepts, and proposes that the two problematic logical expressions can in fact be given ready interpretations that are neither necessarily true nor necessarily false.

3.3. Psychological supervaluations

The threshold model can give a straightforward account of the falsity of $A(x)\&\neg A(x)$, through a more detailed consideration of the nature of gradedness. According to this model there are three major places where variance may occur in the process of categorization across different occasions and by different individuals (see Barsalou, 1987; Hampton, 1995). The representation of the instance x may vary, as may both the representation of the concept A and the threshold value of similarity required for categorization. Each of these variance components may then be attributed to any of three sources (and their interactions): random variation in time within an individual; consistent individual differences; and contextual effects (broadly conceived). But if we assume that in any particular act of categorization, all three sources of variance are held constant (that is, the categorization is made relative to a particular individual's representation, similarity metric and threshold placement on a particular occasion,) then of course an instance cannot be both above the threshold and below the threshold at the same time. It therefore follows that in each individual act of categorization, $A(x)\&\neg A(x)$ will be false. Hence it will be necessarily false in every case. All respondents, when asked to evaluate

A(x)&¬A(x) will consider the question of whether x is an A from their own viewpoint in the particular context in which the question is asked, and so all will give opposite valuations to $A(x)$ and to $\neg A(x)$.

The same account handles *A(x)OR¬A(x)*, since for any given instance + individual + context occasion of categorization, if an instance is not above the threshold then it must be below. Hence it will be true in each individual case that the instance is either sufficiently similar or not sufficiently similar to the category to be a category member. A similar proposal, relating categorization to perspectival context, has been made by Braisby (2005). Note that if categorization is not tied to a particular person and context (or to some fixed group of people and contexts) then it would follow that *A(x)&¬A(x)* could be paraphrased as "x is an A for some person (or group) $p_1$at some time t_1, and x is not an A for some other person (or group) p_2 at some other time t_2." This paraphrase would seem to make the contradiction always true for all x that lie in the borderline region of a concept (that is for any x where some person at some time would consider it to belong and some other person at some other time would consider it not to belong), which is hardly the desired result.

The psychological supervaluation account also offers an answer to another difficult case for fuzzy logic. If two items occupy identical positions on a scale (for example two men, John and James are of identical height), then the proposition "John is tall and James is not tall" should always be false. This result could be achieved if it assumed that people understand that for a particular person on a particular occasion, the act of categorization could not produce different answers given the same input value, and so John and James cannot have different values for category membership. The case of identity is another constraint on possible completions, like the monotonicity constraint from which K&P derive a scale for M.

Psychological supervaluation shares a considerable similarity with K&P's supervaluations. Just as supervaluation applies across the set of completions, so for psychological supervaluation people consider the set of individual acts of categorization and then apply a supervaluation across such acts. There are however the following important differences from K&P. First, there is no requirement to distinguish the borderline region as a clearly defined range of similarity, since it is assumed that the threshold placement and other sources of variance give rise to a continuous threshold function. We therefore avoid the difficult second order vagueness problem of deciding, for example, the exact range to the right of centre on the piano keyboard for which the issue of a note's being high or not is in question. Second, whereas K&P's account is a purely logical account which generates a degree of membership for all the logically possible completions of the partial model, the account offered here is a psychological model, in which the variable categorization behaviour of an individual is explained in terms of a stochastic model based on underlying similarity. The intuition of logical necessity is derived from supervaluating over a set of categorization acts for specific individuals at specific times, rather than supervaluating over the set of potential category members.

The account also serves to underline the point that the Threshold model, in common with most prototype models and contrary to some other accounts of vague concepts (and some misconceptions of prototype theory itself), does in a technical sense provide "a necessary and sufficient condition for category membership." An item is a member of a category if and only if it has a similarity to the prototype that is above the threshold level. Given a correct account of similarity, such a categorization rule could then be made explicit through a complex Boolean

function of the relevant concept features, although it would hardly explicate the meaning of the concept. The more important point, of course, is that unlike traditional criteria, these conditions are liable to be slightly different for each individual on each occasion of use.

Where psychological supervaluation fails is in accounting for individual intuitions of vagueness. By this account, on any particular occasion when an individual makes a categorization decision there would be nothing to indicate whether this was a clear-cut case or a borderline case subject to revision. Yet individuals do seem able to distinguish definite decisions from borderline decisions (Kalish, 1995). In fact the account requires that each individual have knowledge of the likely variability in categorization acts performed by others. Section 4 returns to this question.

3.4. Combining prototypes

A different answer to the question of how people understand $A(x)\&\neg A(x)$, or $A(x)OR\neg A(x)$ involves examination of the results of research on the logic of conceptual combination, and in particular research into how conjunction, disjunction, and negation of concepts affects measured category membership judgments. The solutions offered in the previous two sections stayed true to the common shared intuition, owed to Aristotle and embodied in most logics, that a proposition must be either true or false once all epistemological uncertainty has been resolved (the Principle of Bivalence). However it is also interesting to see if an account can be given of the alternative intuition that the flexibility of our concepts allows us to say interesting and meaningful things about the world that on a strict view would be considered logical contradictions. The interpretation of statements such as "a tomato is both a fruit and not a fruit" as meaning that it is somewhere between being a fruit and not being a fruit has been dismissed as a pragmatically driven interpretation of a loosely phrased sentence (Osherson & Smith, 1981; 1982). In the next section, by looking at data on (a) how prototypes change when they are negated and (b) how prototypes combine in conjunctions or disjunctions, the source of such interpretations is revealed.

3.4.1. Conjunctions

Under what circumstances might one meaningfully assert the truth of a statement of the form $A(x)\&\neg A(x)$? The answer lies in an individual's ability to reflect on the range of variability in the conceptual representations of x and A. If $A(x)$ and $\neg A(x)$ express different truths, arising from different ways of interpreting the two terms, then one can believe in both without self-contradiction. As outlined above, the threshold model proposes that variable categorization across individuals and occasions may arise from variance in instance representation, variance in category concept representation and variance in the placement of the threshold. Any of these three sources of variance can also be used to provide an interpretation for $A(x)\&\neg A(x)$. Let us consider them in turn.

First, an interpretation may be given through a change in the representation of the instance x being classified. Thus in saying "Chess is both a sport and not a sport," one might think of two instantiations of chess. One kind of chess is highly skilled and competitive and has a World Championship providing entertainment to an audience of fans and is reasonably clearly a sport. In another guise chess can be thought of as a board game like checkers that you can

play with your children as a form of pleasant pastime, and so is reasonably clearly not a sport. The instance class may be considered as disjunctive, so that "x is both an A and not an A" becomes "Some forms of x are A, whereas other forms are not A." In logic, the universally quantified "All x are A and not A" is not of course equivalent to the particular "Some x are A and some x are not A," but given that the first logical interpretation is clearly false, in contexts where its utterance would break the Gricean maxims of quality and relation, (Grice, 1975) then the second logical formula may offer a quite reasonable interpretation of the speaker's intended meaning, and hence a meaning that a speaker could choose to intend. Hampton (1982) showed that it is quite common for people to accept the truth of a statement, even when they accept that there exist counterexamples. People will readily agree that a chair is a type of furniture, while also accepting that car seats and chairlifts constitute counterexamples to a universally true "all chairs are furniture." Even explicitly marked subsets such as "school furniture" or "furniture that is a household appliance" are not actually treated as strict subsets of furniture (Hampton, 1988b). In a similar way, people will consider "flies" to be a centrally important feature of birds, in spite of their knowledge of counterexamples such as penguins and ostriches.

Second, an interpretation may invoke variability in the representation of the category concept A. Rather than finding two instantiations of chess, we may retrieve two different instantiations of the notion of sport. Thus we might think "Chess is a sport in the sense of sport being a competitive activity which is enjoyed, requires skill and has a national governing body, but chess is not a sport in the sense of sport as a physical activity that involves exertion, sweat and physical skills." Hence x is an A if more attention is paid to some subset of the attributes of As, while x is not an A if attention is paid to some other subset. The difference here is that we are not focussing attention on the disjunctive nature of the instance class—forms of chess—but rather on the polymorphous nature of the category sport itself. Unlike the first interpretation, this category-based variance would provide an account of how $A(x)\&\neg A(x)$ may be non-zero for both classes of instances and individuals within those classes.

Finally, a third interpretation could involve recognition that the threshold on similarity required for categorization may shift as a function of either individuals–"Chess is a sport according to people who take a broad view, and not a sport according to those who take a narrow view"—or contexts "I would call chess a sport in a broad sense, but not a sport in a narrow sense." The notion of concepts being interpretable in broad and narrow senses is very familiar, and is reflected in the English language in "hedges" such as "broadly speaking" versus "strictly speaking" (Lakoff, 1973; Malt, 1990). Some of these hedges appear to involve a shift in dimensional weights (Hampton, 1998), but it is plausible to suppose that they may also often be used to describe different threshold levels. Note that this third interpretation involves people's meta-beliefs about the usage of terms in their language. They can reflect on how some people would say yes, and others no, or even on how their own decision would change depending on the context or perspective taken.

What evidence is there that people interpret conjunctions in a way that is non-intersective? Hampton (1997a) examined this question. To take an illustrative example, six independent groups of participants provided judgments of category membership for a list of items such as castle, phone-box or tent, for the following six categories:

1. Buildings
2. Dwellings
3. Buildings that are dwellings
4. Dwellings that are buildings
5. Buildings that are not dwellings
6. Dwellings that are not buildings

Intersective interpretations of the conjunctions would predict that (if categorizers are assumed to be fully rational) the probability of being categorized in $A\&B$ (groups 3 and 4) or in $A\&\neg B$ (groups 5 and 6) should be constrained within limits set by the probability of being categorized as A and as B separately (groups 1 and 2). However the data did not respect these limits. Items were frequently categorized in the complex concepts with greater probability than should have occurred on the basis of the simple concepts alone. Analysis of the individual items showing this effect suggested that the representation of categories was undergoing just the kinds of variation that are implicated in the interpretations of $A(x)\&\neg A(x)$ described before. For detailed accounts of these effects in terms of a model of default attribute inheritance see Hampton (1987, 1997a).

For the present purposes, note that the effects of conjunction and negation on prototype concepts often only approximate the equivalent operations in Boolean logic. This evidence then licenses us to consider that similar effects may be found with the $A(x)\&\neg A(x)$ case. According to Hampton (1987) when concepts are conjoined, a process of conflict resolution operates on those default attributes of the two concepts that are incompatible with each other in order to generate a composite prototype to represent the conjunction (see also Thagard, 1997). For example pets live at home and are warm and furry, while fish live in the ocean and are cold and slimy, yet pet fish live at home like pets, but are cold and slimy like fish. In each case, one dimensional value wins out over the other in determining the prototype for a pet fish.

Clearly when conjoining A and $\neg A$ there will be many such conflicting attributes. The more central an attribute is for A, the more likely it is to be negated or simply absent for $\neg A$. The resolution of the conflict can be achieved through allowing the composite to inherit either one or the other of the attribute values.

An illustration will make this clearer. Take the example of the self-contradictory conjunction "A dwelling that is not a dwelling." Let us suppose, as is reasonable, that two central attributes of dwellings are *is lived in* and *was intended to be lived in*. When forming the composite for this conjunction we may find that NOT A DWELLING would normally imply the attributes *not lived in* and *not intended to be lived in*. (These would appear to be reasonable inferences to draw from a statement that some location was not a dwelling—see Hampton, 1997a for data on the effects of negation on the attributes and category membership of conjunctions.) The conjunction therefore has (at least) these two conflicting attributes to resolve. In order to reflect both constituents, suppose that the composite takes one attribute from each. We would therefore have two possible composites, containing (among other things) the following attributes:

lived in + not intended to be lived in
not lived in + intended to be lived in (11)

The first composite (11) would match an example such as "an unsanctioned home, an office with a bed in it where someone is illicitly staying the night." Alternatively, the second (12) would match the example "an empty house–created as a dwelling, but nobody lives there." It can therefore be seen how the process of combining default attributes, and selecting one over the other in the case of conflict, leads to intuitively plausible interpretations of the $A(x)\&\neg A(x)$ contradiction.

The same analysis can be applied in many other cases, so that for example "A weapon that is not a weapon" could be either something from another category being used as a weapon (a kitchen knife), or it could be something intended as a weapon that no longer can function as one (a cannon on a war memorial). The prediction here would be that whenever a person can offer answers to the following two questions:

Why might you categorize this as an X?
Why might you not categorize this as an X?

then (assuming that there was no further relevant information that was not known about the object) they should be willing to agree that the object was to some appreciable degree an X which is not an X.

To conclude the discussion of conjunctions, I have argued that O&S may be right when they state that there are pragmatic interpretations available for "loosely phrased sentences." My emphasis however has been to explicate these interpretations through empirically grounded models of how prototype concepts are conjoined to form complex concepts. Rather than dismissing them as being an aberration on the logical landscape, I have argued that they may play a major role in our conceptual thought.

3.4.2. Disjunctions

In a similar way, one can look at the problem of the disjunction of $A(x)$ and $\neg A(x)$, which under the normal rules of Boolean logic should be universally true. Evidence on the ways in which disjunctions of prototype concepts are formed can be found in Hampton (1988a), where the pattern of quasi-logical combination was repeated. Among other category pairs, participants were required to judge whether items such as Apple, Spinach or Mushroom were in each of two separate categories (Fruit, Vegetable) and then whether they were in the disjunction of the two (Fruit or Vegetable). The item Mushroom, which was never considered to be a Fruit, and was only considered to be a Vegetable by 60% of participants, was considered to be in the class "Fruits or Vegetables" by 100% of participants. In other words at least 40% of the participants were not applying Boolean logic to the problem, (assuming that they were applying a consistent criterion of confidence for saying "Yes" in the different tasks). The results could not be explained by lack of knowledge in this case (it is quite possible to know that something is in category A or category B but not which), because all participants agreed that mushrooms were NOT fruit. There was no uncertainty about that. They should then have been equally uncertain about whether mushrooms were vegetables or whether they were fruit or vegetables.

Speculatively, one can propose that because the two concepts have many common features (*plant, eaten*, etc.) and many alignable differences (*grow on trees* vs. *grow on ground*, *eaten raw* vs. *eaten cooked*) the two prototypes can be combined into a single more general concept

with more broadly defined features (*grow, eaten*) corresponding perhaps to the US concept of "fresh produce," or "greengroceries" in the UK. Alignable differences are then just omitted from the generalised concept.

Turning to the logical problem for disjunctions of $A(x)OR\neg A(x)$, the intensions of A and $\neg A$ will of course be perfectly aligned, with more or less identical dimensions. Hampton (1997a) found that when a category is negated within a particular domain, domain general attributes are unaffected, while more specific attributes are either lost all together, or are themselves negated. Thus Dwelling was "a place where people live with doors, windows and heating" while Not a Dwelling was "a place where people do not live."

The disjunctive combination of the concepts A and $\neg A$ would therefore proceed as follows. Any attributes (such as "is a place") that remain positive in $\neg A$ will be inherited in the composite unaffected, while attributes which are negated in $\neg A$ ("is not lived in") will be disjoined with their un-negated form, so that the dimension will be dropped from the representation. Hence $A(x)OR\neg A(x)$ ends up as a more superordinate domain level category ("is a place"). It follows that the degree to which an instance x belongs in the disjunctive category $AOR\neg A$ will be equal to the degree to which it belongs in the domain of which A is a subset. The prediction therefore follows that an item that clearly falls within the correct domain should receive a higher categorization than one that does not. The category disjunction "Either a fruit or not a fruit," should apply very obviously to an item like a tomato, but less so to an item like a lamb chop, and not at all to an umbrella.

As in the case of conjunctions, by examining the processes of alignment and combination of concept representations, a broader account can be given of the logical formula of the disjunction of a category predicate and its negation. In particular it has been argued that for vague concepts, $A(x)OR\neg A(x)$ can be given an interpretation that assigns a truth value that is not always 1, but reflects the truth of the degree of membership of x in the superordinate domain in which A lies.

4. Conceptions of vagueness

The discussion to this point has located the problem of vagueness as occurring through the variability of conceptual representations and thresholds across individuals and contexts. (Recall that vagueness refers to the gradedness of membership M rather than typicality T although it has been argued that both notions are indices of the same underlying measure.) I have proposed that a particular implementation of prototype theory—the threshold model—can account for the different phenomena of degrees of membership M and typicality T without the need for separate accounts. I have offered three accounts of how the model can handle self-contradiction and tautology, through explicit marking of logical form in certain syntactic forms, through a psychologically based version of supervaluations and through combining prototypes to provide interpretations that are not analytically true or false.

But no account has yet been given of why most people have the strong intuition that the application of vague terms does not involve binary truth conditions. For example the psychological supervaluation account may be proved wrong if, even when I am situated in a defined context at a given moment in time, I am still unwilling to clearly categorize an

instance within a particular category, in spite of having a threshold criterion which allows me to make a binary judgment (Kalish, 1995). I may be faced with a particular tomato in a particular situation and be asked if it is a fruit, understood in a certain way dictated by the communicative context, and I may still want to say "I can't really say yes or no, it is to some extent, but not fully." The existence of these intuitions was shown in a study (Hampton, 2004) in which people were first given borderline cases of categories and asked to say whether they belonged using three options "Clearly" "Intermediate" and "Clearly not." If they chose the intermediate response they were then asked to choose a reason for doing so from a range of possibilities including ignorance of the case or the category, ambiguity of the terms, or lack of precision in whether a broad or narrow sense was intended for the category. Reasons chosen were evenly divided between these three options, with ignorance largely being used for biological categories (birds, fruit etc.). There is therefore evidence that people are aware of the fuzziness of category boundaries, and they don't attribute this vagueness just to ignorance.

To provide an account of these intuitions, it is necessary to take a broader perspective on what the conceptual system has evolved to do.

As Rosch and many others have noted, we use our conceptual system to reduce the complexity of our environment to a manageable level by focussing attention on relevant dimensions. Inevitably this data reduction process will introduce levels of simplification and approximation. Lakoff (1987) refers to this process as idealization. Where it is useful or important to our purposes, we may choose to take a non-vague approach. We define concepts like grandmother, dollar or president by explicit stipulation. Concepts that have a role to play in the economic, legal or social life of a community tend to be given explicit verbal definitions that people can use to justify and legitimate decisions. But most of the time we are covering up the fact that the world does not have simple joints where it can easily be carved (Kosko, 1990). As Wittgenstein (1953) noted it is possible to learn the clear cases of a concept without knowing how to determine the boundaries of its conceptual category.

This difficulty in fixing borderlines leads to a strategy of ignoring conceptual category boundaries in favor of learning conceptual cores (be they theories, prototypes or representative exemplars). The boundaries remain fluid for good reasons. When the world changes, or we discover new facts about it, our concepts can adapt to the change while their identity is still tracked. When context requires a different set of dimensional weights, the closer an instance is to the centre of the concept cluster, the less it will change. Centres are also easier to teach by ostension, since there are few if any alternative categories to which an item will belong, within a particular categorization scheme. This focus on central cases also explains why we have so little insight into the way dimensions determine category membership. As well as showing vagueness and typicality, our concepts also display opacity and genericity (Hampton, 2006). Opacity, as used in this context, is our inability to introspect and give a clear account of the content of a concept. A small number of concepts are not opaque (for example UNCLE or WIFE), but most are. Genericity is the fact that most properties that we would associate with a concept tend to be true "generally speaking" or "as a rule," rather than necessarily or universally true. Both opacity and genericity are clear evidence that we learn concepts by learning about central cases and not about boundary cases.

How do we live with vagueness? Vagueness in the use of a category label (and hence in seeing an object as of a particular type) is constrained by socialization. Following Grice

(1975), there is a powerful motive, even a civic duty, to use words in a way that our interlocutor will understand. We are members of a linguistic community (or several) and we learn and note the usage of terms by hearing and reading others using them. Dictionaries codify current accepted usages, and are useful as arbiters of meaning, but in the long run they follow changes in language use, and act merely as a regulator to stabilize the continuing process of linguistic evolution. (The belief that dictionaries provide normative/prescriptive meanings for words is however still strongly held in many quarters). Word meanings (and hence the concepts that the words convey) are learned largely through observation of word use in context, and this word use will predominantly occur in prototypical situations. Hence the use of concept terms at the borderline of a class is not often observed in the environment, and so may remain vague or undefined. In fact objects that are atypical of any category will tend not to have a consistent term applied to them at all.

Given this view of language terms as having their meanings based in social use, how can one fix a notion of truth? There are many interpretations of the notion of fuzzy truth. Williamson (1994), for example, supports the epistemic or "vagueness as ignorance" interpretation of vague statements. According to this hypothesis, the truth conditions of a statement such as "a man who is 1.70 m high is tall" are of the same kind as those of a statement such as "a man who is 1.70 m high is above average height for a 30-year-old male." The vagueness of the first is considered as equivalent to the epistemological uncertainty in judging the second. The application of the term "tall" is only vague because we do not know the correct criterion that divides the well-defined class of tall men from that of not-tall men, in just the same way that we do not know the precise average height of 30-year-old males.

Interestingly, this hypothesis shares much in common with the hypothesis of psychological essentialism proposed by Medin and Ortony (1989), in which intuitions of the fuzziness of categorization are considered to be the result of people's belief that there is some defining essence of the category, of which they are sadly ignorant.

Although there may be good logical grounds for liking the hypothesis, there is little empirical evidence for the Vagueness as Ignorance thesis. Bonini et al. (1999) showed that both kinds of judgment (vague ones and uncertain ones) show "truth gaps," in that the minimum height that is considered tall (or above average) is greater than the maximum height that is considered not tall (or below average). But this is a weak argument for the thesis that vagueness just is uncertainty, since there are many other ways in which the two judgments differ. More seriously, the hypothesis has major problems in other respects. It is committed to there being an objective answer to the question "How tall is tall?" that is exact—in the same way that the average height of a well defined population at any moment in time is exact in reality, even if in practice it would be impossible to calculate it to more than a certain level of precision. But there is no way in which one could explain just how the word "tall" came to be even approximately associated with this precise objective external value. Likewise there is no answer to the question of whether there is one single value for TALL for all speakers of English, or whether different dialects (for example, Scottish versus New Zealand) would each have their own value—perhaps reflecting different mean heights of different cultures. If there is just one objective value, then one would need a host of explanations for why whole cultures have a systematic bias to get the usage wrong (assuming that for a significant number of vague terms one could show cultural differences between usage). If there is more than one

objectively true value, then there is nothing to prevent us asking the same question recursively at increasing levels of specificity until finally we are led to the notion that since each of us speaks our own idiolect, we each of us have a different precise objective concept of TALL to which our own concept is linked.[5]

People are clearly ignorant of many things, and believe many more–but the belief that one is ignorant of the precise criterion for application of a term does not give validity to the existence of that criterion. Furthermore evidence for essentialist beliefs has not been unequivocal even for natural kinds (Hampton, 1995; Kalish, 1995; Braisby, Franks & Hampton, 1996; Hampton, Estes, & Simmons, 2007). What people may be estimating when giving a judgment of degree of membership M is how comfortable they would feel using the term in a certain way or context, and this sense of easiness will be more or less directly related to the proportion of language users who would agree to the use of the word in that context (Black, 1937). It is in this sense that it may be reasonable to treat probability of categorization as a measure of graded membership M.

The proposal here is that intuitions of vagueness are based on beliefs about language use in one's language community, and meta-beliefs about the variability of one's own conceptual representations across contexts and occasions (Barsalou, 1987). We base our intuitions on our past experience with linguistic usage of a term, as it is reflected in the prototypes we associate with each word. Because conformity to the social norm is imperfect, there is plenty of room for people to arrive at different answers and to be inconsistent with their own answers on different occasions. There is a dynamic equilibrium between forces leading to identical usage of terms, and those leading to idiosyncratic usage (Honkela & Winter, 2003). Usages may be thought of as competing for survival in a Darwinian selection process (Sperber & Hirschfeld, 2004).

If the intuition of vagueness has to do with first trying to assess the degree to which others would categorize the instance in the category, and then realising that there is no strong consensus, this process is of course reciprocal. That is, the others, whose usage we are estimating, are also uncertain about the correct usage, since they too are basing their own use on what they feel is the general consensus. In this sense we are not just all ignorant of the normatively correct usage. There is no such thing since the consensus is arrived at by the aggregation of everyone's different sense of uncertainty.

Much like estimating a subjective probability, trying to use a term correctly involves assessing a relative frequency, and heuristics such as availability and representativeness may be involved in making the assessment (Tversky & Kahneman, 1974). Representativeness is, however, just the same notion as typicality T. So people judge that the more similar an object is to typical cases of a category, the more likely it is that they themselves and others will use the category label to describe it. Thus T is used as a means of assessing M (with the proviso that the range of the functions is different as discussed in the earlier section).

Individuals might even apply supervaluations here, by treating other members of the language community, or different contexts of use, as completions of a partial model. In other words the set of possible completions that figure in the supervaluationist account could be thought of as corresponding to the set of possible usages or beliefs in the population of language speakers (or even the set of speakers themselves). One could then use the logic of supervaluations to conclude that the probability of someone believing x is a fruit and the same person on the same occasion believing that x is not a fruit is zero, and similarly that

the probability of some person on a particular occasion either believing that x is a fruit or believing that x is not a fruit is 1 (assuming that all relevant information about x was available). Similarly one could sensibly believe that there are constraints on individual usage such as that the probability that someone believes a man of 1.80 m is tall and the same person on the same occasion believes that another man of 1.80 or 1.82 m is not tall is zero. Applied to meta-beliefs about how other speakers might use a term, it is possible for the psychological supervaluation account to handle both the logic of self-contradiction and tautology and also to explain intuitions of degrees of membership. There may also be a normative element to the judgment. A person may believe it quite possible that a categorizer would arbitrarily call one person of 1.80 m tall and another of identical height not tall on the same occasion, yet believe that to do so would be incorrect or unjustifiable. We therefore may be applying a naïve notion of rationality when we consider our metabeliefs about how people should categorize in order to draw supervaluationist conclusions.[6]

5. Conclusion

In conclusion, categorization involves a comparison of a represented stimulus with some internalised representation of a class. Typicality T and graded membership M can be each derived from the same underlying measure of similarity, and the problems that Zadeh's fuzzy logic had deducing that $A(x)\&\neg A(x) = 0$, and that $A(x)OR\neg A(x) = 1$, can be captured either by appeal to the explicit logical form, or by assuming a psychological version of K&P's supervaluation thesis. Alternatively a model of how prototypes combine may be used to provide an account of the variety of interpretations that may be placed on such expressions in a Gricean context. Actual gradedness of categorization as seen in the variability of categorization behavior can be explained in terms of the same underlying measure of similarity as can typicality. However people's intuitions of gradedness are based on meta-beliefs about the use of terms and concepts within one's language community. Vagueness is the inevitable result of a knowledge system that stores the centres rather than the boundaries of conceptual categories.

Acknowledgments

The author wishes to acknowledge the following for their help in clarifying the issues addressed in this paper: Zachary Estes, Martin Jönsson, Hans Kamp, Gregory Murphy, Claire Simmons, Steven Sloman, Timothy Williamson, and members of the London Concepts Group. Three anonymous reviewers also provided invaluable help in identifying errors and inaccuracies in the text. Any vagueness that remains is entirely of my own devising.

Notes

1. Probability of categorization will reflect M but will also reflect other sources of individual variability such as ambiguity of the terms, ambiguity of context, or variation in people's conceptual beliefs.

2. Note that not all dimensions increase typicality indefinitely. One could argue for example that someone may become so violent and hateful that the word "bully" is no longer an adequate description, and another term like "psychopath" may take over. Only the concept that lies at the end of a series such as LARGE, ENORMOUS, HUMUNGOUS may have a function for *T* that increases indefinitely.
3. This function ensures continuity for the first derivative but not the second. Should continuity for all derivatives be required a function based on a sine curve could be used– and should the function need to be asymmetrical then a suitable nonlinear transform could be applied to the sine, such as raising it to some suitable power. The aim of the function is simply to illustrate that there are many simple functions which would have the appropriate properties.
4. In fact before the thermodynamic definition of temperature provided a theoretically grounded account of the construct, temperature scales had to specify the instrument used for measurement, since the thermic expansion properties of alcohol, metal or mercury were not equivalent, and there was no principled way to decide which should be used as the basis for calibrating the others.
5. The problem of every individual having slightly different concepts is one that will apply to most psychologically realistic descriptivist models of concept representation. The question of how they all manage to be representations of the same concept is an important one that would take us too far away from the issues considered here. For the present, I will just note that people frequently do have different concepts—they use terms differently to categorize the world, and associate different properties with conceptual classes.
6. It may appear that the appeal to meta-beliefs and the pragmatics of word-use in this final section is at odds with the context-independent treatment of degrees of membership in the earlier sections of the paper. I do not intend the two approaches to be incompatible. I believe that we do possess stable default representations of concepts in long-term memory from which the meaning of a phrase in context is constructed. The pragmatics has to have a stable base on which to work. The frequent use of a term in particular pragmatic settings however will have a corresponding influence on the content of the stored representation of meaning of the term in memory.

References

Ahn, W. K., & Dennis, M. J. (2001). Dissociation between categorization and similarity judgements: Differential effect of causal status on feature weights. In U. Hahn, & M. Ramscar (Eds.), *Similarity and categorisation* (pp. 87–107). Cambridge: Cambridge University Press.

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.

Black, M. (1937). Vagueness: an exercise in logical analysis. *Philosophy of Science, 4,* 427–455.
Bonini, N., Osherson, D. N., Viale, R., & Williamson, T. (1999). On the psychology of vague predicates. *Mind and Language*. 14, 377–393.
Braisby, N. (2005). Similarity and categorization: Getting dissociations in perspective. In K. Forbus, D. Gentuer, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 150–155). Mahwah, NJ: Erlbaum.
Braisby, N., Franks, B., & Hampton, J. A. (1996). Essentialism, Word Use, and Concepts. *Cognition, 59,* 247–274.
Cohen, B., & Murphy, G. L. (1984). Models of concepts. *Cognitive Science, 8,* 27–58.
Evans, J. St. B., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11,* 295–306.
Fine, K. (1975). Vagueness, truth and logic. *Synthese, 30,* 265–300.
Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Clarendon Press.
Fodor, J. A., & Lepore, E. (1996). The pet fish and the red herring: Why concepts aren't prototypes. *Cognition, 58,* 243–276.
Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, Vol. 3* (pp. 41–58). New York: Academic Press.
Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18,* 441–461.
Hampton, J. A. (1982). A demonstration of intransitivity in natural categories. *Cognition, 12*, 151–164.
Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15,* 55–71.
Hampton, J. A. (1988a). Disjunction of natural concepts. *Memory & Cognition, 16,* 579–591.
Hampton, J. A. (1988b). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition, 14,* 12–32.
Hampton, J. A. (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). Conceptual combination: Conjunction and negation of natural concepts. *Memory & Cognition, 25,* 888–909.
Hampton, J. A. (1997b). Psychological representation of concepts. In M. A. Conway (Ed.), *Cognitive Models of Memory* (pp. 81–110). Hove: Psychology Press.
Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition, 65,* 137–165.
Hampton, J. A. (2001). The role of similarity in natural categorization. In U. Hahn, & M. Ramscar (Eds.), *Similarity and categorisation* (pp. 13–28). Cambridge: Cambridge University Press.
Hampton, J. A. (2004). Reasons for vagueness. Paper presented to the 45th Annual Meeting of the Psychonomic Society, Minneapolis, November.
Hampton, J. A. (2006). Concepts as Prototypes. In B. H. Ross, (Ed.), *The Psychology of learning and motivation: Advances in research and theory, 46* (pp. 79–113). New York: Academic Press.
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., Estes, Z., & Simmons, S. (2007). Metamorphosis: Essence, appearance, and behavior in the categorization of natural kinds. *Memory and Cognition* (in press).
Hampton, J. A., Estes, Z., & Simmons, C. L. (2005). Similarity-based contrast in color judgment. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31,* 1459–1476.
Hampton, J. A., Dubois, D., & Yeh, W. (2007). The effects of classification context on categorization in natural categories. *Memory & Cognition*, *34*, 1431–1443.
Honkela, T., & Winter, J. (2003). Simulating language learning in community of agents using self-organizing maps. Helsinki University of Technology, Publications in Computer and Information Science, Report A71.
Jönsson, M. L., & Hampton, J. A. (2006). The inverse conjunction fallacy. *Journal of Memory and Language, 55,* 317–334.
Kalish, C. W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory and Cognition, 23,* 335–353.

Kamp, H. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal Semantics of Natural Language* (pp. 123–155). Cambridge: Cambridge University Press.
Kamp. H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition*, 57, 129–191.
Keefe, R., & Smith, P. (1997). Theories of vagueness. In R. Keefe, & P. Smith (Eds.), *Vagueness: A Reader* (pp. 1–57). Cambridge: MIT Press.
Kosko, B. (1990). Fuzziness vs. probability. *International Journal of General Systems, 17,* 211–240.
Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2,* 458–508.
Lakoff, G. (1987). *Women, fire and dangerous things*. Chicago: Chicago University Press.
Malt, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language, 29*, 289–315.
McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory & Cognition, 6*, 462–472.
Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.
Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical Reasoning* (pp. 179–195). Cambridge: Cambridge University Press.
Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.
Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 700–708.
Osherson, D. 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., & Smith, E. E. (1997). On typicality and vagueness. *Cognition, 64,* 189–206.
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.
Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192–232.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22.
Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology, 35,* 1–33.
Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science*, 12, 485–527.
Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences, 8,* 40–46.
Storms, G., De Boeck, P., van Mechelen, I., & Ruts, W. (1996). The dominance effect in concept conjunctions: Generality and interaction aspects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1266–1280.
Thagard, P. (1997). Conceptual combination, Coherence and Creativity In T. B. Ward, S. M. Smith, & J. Viad, (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 83–110). Washington DC: American Psychological Association Press.
Thibaut, J. P., Dupont, M., & Anselme, P. (2002). Dissociations between categorization and similarity judgments as a result of learning feature distributions. *Memory & Cognition, 30,* 647–656.
Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.
Williamson, T. (1994). *Vagueness*. London: Routledge.
Wittgenstein, L. (1953). *Philosophical Investigations*. New York: Macmillan.
Zadeh, L. (1965). Fuzzy sets. *Information and control, 8,* 338–353.