ELSEVIER

Cognition 64 (1997) 189–206

COGNITION

Discussion

On typicality and vagueness

Daniel Osherson[a,*], Edward E. Smith[b]

[a]Rice University, Houston, TX, USA
[b]University of Michigan, Michigan, USA

In their meticulous article entitled ''Prototype theory and compositionality,'' Kamp and Partee (henceforth K&P) analyze the phenomena of vagueness and prototypicality, and offer a solution to one aspect of the compositionality problem for imprecise concepts. The purpose of the present rejoinder is to mark our agreements and disagreements with K&P, as well as to comment more generally on the issues surrounding typicality and degree of membership. Although we disagree with K&P on fundamental issues, there can be no doubt that their article brings new clarity and fresh perspectives to a complex topic.

We begin by characterizing the phenomena of typicality and graded membership as we conceive them. In what follows, unaccompanied page references are to Kamp and Partee (1995). To refer to Osherson and Smith, K&P use O&S.

1. Two kinds of gradedness

Discussion of prototypes and compositionality involves two types of gradedness, now considered in turn.

1.1. Typicality

Typicality refers to the extent to which objects are 'good examples' of concepts. To illustrate, let Q be Queen Elizabeth's throne, and let I be the most popular chair sold by the furniture chain IKEA. Then, the intuitions of many people are that I is a better example of *chair* than is Q in the sense of better serving to communicate the concept in question. Thus, we say that I is 'more typical' than Q with respect to *chair*. Typicality is related to a variety of performance measures in classification (see Smith and Medin, 1981; Hampton, 1993) for reviews). However, we follow

* Corresponding author. e-mail: osherson@rice.edu, eesmith@umich.edu

PII S0010-0277(97)00025-5

Rosch (1978, p. 36) in taking goodness of exemplification to be the central variable in typicality.

In the early literature, typicality was conceived as proximity to the mental prototype of a concept (Rosch and Mervis, 1975; Rosch, 1978). However, it may be more accurate to recognize similarity to prototypes as just one potential mechanism for producing intuitions of typicality.[1] Alternative mechanisms apparently exist since there are concepts that lack prototypes while possessing degrees of exemplification. An example (for us) is *building*. No prototype comes reliably to mind even though most banks strike us as better examples than most barns. The possibility of graded typicality in the absence of a prototype is particularly important for complex concepts, which often do not have prototypes (as observed in Fodor (1981, pp. 296–297)). An example is *defective supercomputer*, which likely lacks a prototype for most people. Yet we can still judge the kind of defects that make for good versus bad examples of the concept (thus, a supercomputer that multiplies incorrectly is a better example than one that occasionally dims its monitor).

To quantify typicality with respect to a given concept *con* we use K&P's notation c^p_{con}, denoting a function that maps the domain of discourse into real numbers. Thus, our earlier claim about Q and I may be written as

$$(1) \quad c^p_{chair}(I) > c^p_{chair}(Q).$$

We do not assume, as do K&P, that the range of c^p_{con} constitutes a closed interval like [0,1]. As discussed later, we think it is a central property of typicality that it reaches no maximum value.

1.2. Degrees of membership

Gradedness also arises when we consider the applicability of concepts to objects. Thus, in addition to admitting instances of varying typicality, *chair* is vague inasmuch as it seems to apply in graded fashion to objects whose potential use for single seating is not clear. To measure the graded applicability of a concept *con* to objects, we follow K&P and use the notation c^e_{con}. Unlike for typicality, the range of this function may be taken to be the closed interval [0,1], since applicability is generally thought to reach a maximum or a minimum in unambiguous cases (thus, *chair* applies to I as well as it applies to anything else).

It is widely recognized that almost every term in natural language admits some vagueness. However, this does not imply that clear cases of applicability are rare exceptions. Only a strange form of skepticism would lead us to doubt that many of the objects in our homes are (clearly and unambiguously) chairs, or that 1600 h in the summer is (clearly and unambiguously) daytime. To restrict claims of graded applicability to the genuinely liminal cases, we rely on the following principle.

[1]This point is already expressed in Rosch (1978, p. 38). For a history of research on prototypes, see Lakoff (1987, ch. 2).

(2) Let the sentence formed by applying concept *con* to object o be denoted: $con(o)$. If $con(o)$ is plainly true, then $c^e_{con}(o)=1$, and if $con(o)$ plainly false, then $c^e_{con}(o)=0$.

Vagueness is a different phenomenon from typicality, as K&P aptly note on several occasions (e.g., pp. 133, 182). We tried to put the point succinctly in Smith and Osherson (1988, p. 52):

> To appreciate the distinction between typicality and vagueness, note that ... while most of us believe that penguins are atypical of birds, few of us doubt that they are in fact birds; typicality is in play here, not vagueness.

K&P apparently agree with the example since they use it themselves (p. 132).[2] The distinction can also be appreciated by reconsidering the chairs I and Q. Whereas I is a more typical chair than Q many people would agree that *chair* applies perfectly and equally well to both of them. Indeed, the *American Heritage Dictionary* defines 'throne' as 'the chair occupied by a sovereign'. In terms of c^e_{chair}, this yields the following contrast with (1).

(3) $c^e_{chair}(I)=c^e_{chair}(Q)=1.$

2. The K&P theses

In view of the clear distinction between typicality and graded membership, the theses defended by K&P are bold and interesting. We understand them to be the following.

(4) (a) There are simple concepts *con* like *red*, *chair*, and *bully* for which $c^p_{con}=c^e_{con}$.
(b) There are complex concepts *AN* like *striped apple* for which $c^p_{AN}=c^e_{AN}$.
(c) For complex concepts *AN* like *striped apple*, the value of $c^p_{AN}(o)$ is governed by the 'recalibration' principle, discussed in Section 6.

The theses are reflected in K&P's recalibration account of *striped apple* (pp. 163–167), and also in such passages as the following (p. 169).

> As early as section 2, we distinguished between the two c-functions: c^e, which specifies the degree of membership, and c^p, which specifies degree of resemblance to the prototype...
>
> We believe, however, that there are also concepts for which [the] two functions do coincide – concepts for which degree of membership is a matter

[2] The point is also made in Zadeh (1982), Armstrong et al. (1983), Williamson (1994, p. 291), and Kalish (1995). Discussing a class of objects, A, Zadeh says (p. 293): '... an object may be far from the prototype in terms of a given metric and yet have full membership in A'.

> of resemblance to the prototype. We think this is so, for instance, for color concepts such as *red*, for many artifact concepts, such as, say, *chair*, and for such character or personality concepts as, for example, *bully*: for all these concepts there is a genuine sense in which falling under the concept is a matter of resemblance to the cases which speakers conceive as prototypical or central.
>
> We think, moreover, that there is such a logical connection between prototype and membership not only for certain simple concepts, but also for a range of complex ones. ... To see whether this is indeed so let us look once more at striped apple – a concept of which we think it plausible that its membership is connected with prototypicality.

We have doubts about all three theses. Thesis (4a) is examined in the next section. Then we consider (4c), followed by (4b).

3. Typicality versus membership in simple concepts: Thesis (4a)

Thesis (4a) is an existential claim, so it cannot be disproved without exhaustive examination of the class of simple concepts (which we take to be roughly coextensive with concepts given monolexemic expression in English). Although we do not propose to undertake any such examination, it is instructive to consider the concepts advanced by K&P as illustration of their thesis, namely, *red*, *chair*, and *bully*.

Regarding *chair*, we believe that most people's intuitions agree with (1) and (3), above. In other words, the IKEA favorite is a better example of *chair* than is the queen's throne, but the latter is unambiguously a chair, ornate and laden though it be. These judgments contradict (4a) with respect to *chair*. The situation seems to be the same for *bully*. Consider two 13-year-olds, Butch and Marvin. Butch beats up 9-year-olds for fun, pushes 8-year-olds off their bicycles, and is the terror of the 7-year-olds. He wears jeans and a leather jacket, and has perfect eyesight. Marvin is just like Butch except that he wears tortoise shell glasses to correct a mild astigmatism. Most people would agree that both are bullies. Hence, Principle (2) yields $c^e_{bully}(\text{Butch}) = c^e_{bully}(\text{Marvin}) = 1$. On the other hand, because of the glasses, Marvin is a less typical example of *bully* than is Butch, so $c^p_{bully}(\text{Marvin}) < c^p_{bully}(\text{Butch})$. It follows that $c^e_{bully} \neq c^p_{bully}$, contradicting (4a) with respect to *bully*.

There remains the prototype of all prototype concepts, *red*. Even here we think it clear that membership and typicality are dissociable. To be convinced, it suffices to compare focal red (the prototype) to any other, distinguishable wavelength that is unambiguously red. That such wavelengths exist is amply documented in Berlin and Kay (1969). They asked their informants to indicate all those color chips 'which he would under any conditions call x [a given color]' (p. 7). When x was red, their American English informant indicated a range of wavelengths (at each of several brightnesses), only a few of which were 'the best, most typical examples of red' (see Berlin and Kay, 1969, Appendix I, p. 119). Informants for many other languages behaved similarly.[3] Choosing any such non-focal, definite red hue h, Principle (2) yields

(5) $c^e_{red}(h) = 1 = c^e_{red}(\text{focal})$.

In contrast, focal red, being the prototype, is by all accounts the best example of the category (see Berlin and Kay, 1969; Rosch, 1973; Kay and McDaniel, 1978). Thus, $c^p_{red}(\text{focal}) > c^p_{red}(h)$, which in conjunction with (5) contradicts (4a) with respect to *red*.

The foregoing cases suggest that typicality and degree of membership never coincide for simple concepts. To find a counterexample for a proposed illustration *con* of (4a), one need merely identify a highly typical example of *con* along with a similar, but eccentric variant. If all goes well, both will be full members of *con* but differ in how well they exemplify it. It remains possible that a simple concept will be discovered that confounds our pessimism. Beyond the examination of individual cases, however, there is another reason to doubt Thesis (4a). Whereas c^e assumes 0 and 1 as minimum and maximum values, c^p sometimes appears to assume neither. Consider *bully* again, and suppose you think that Butch is such that $c^p_{bully}(\text{Butch})$ is maximal. Now meet Knuckles, who is just like Butch except yet more gleeful as he pummels his victims. Then $c^p_{bully}(\text{Knuckles}) > c^p_{bully}(\text{Butch})$, defeating the claim that $c^p_{bully}(\text{Butch})$ is maximal. Similar examples can be constructed against the claim that any particular (angelic) person occupies the minimum value of c^p_{bully}. Degree of membership and degree of typicality thus appear to be scaled differently, which suggests that they are fundamentally different phenomena. In such circumstances we would be surprised to find that c^e and c^p sometimes coincide.

We now turn to complex concepts and Thesis (4c). As a preliminary, the next section discusses some issues about compositionality on which there is no disagreement between K&P and us. Once these matters are set aside, composition via recalibration can be directly addressed.

4. Agreements with K&P about compositionality

4.1. The existence of a compositional mechanism

K&P believe that the facts about prototypicality 'do not support O&S's global pessimism about the possibility of a compositional theory of combination for gradient concepts' (p. 185). However, we do not deny that there are interesting principles for predicting the typicality properties of complex concepts on the basis of the psychological representations of their constituents. In fact, such principles were proposed and experimentally defended in Smith and Osherson (1984), and then again, in yet more detail, in Smith et al. (1988). The only substantive issues are the character of the typicality principles and their relation to graded membership

[3]See Chierchia and McConnell-Ginet (1992, p. 389) for concordant intuitions from semanticists sympathetic to supervaluation theory.

(and ultimately, to purportedly graded truth). Here indeed we are skeptical about the prospects for a theory of concept membership that attempts to account, as well, for typicality intuitions.[4]

Since typicality and membership strike us as unrelated phenomena, we were careful to insist that our model of typicality was not a theory of conceptual combination:

> We doubt that any approach based only on prototype representations can provide a complete account of conceptual combination. ... There is more to a concept than its prototype (Smith and Osherson, 1984, p. 359).
>
> In particular, we will focus on adjective-noun conjunctions such as *striped apple* and *not very red fruit*, and specify how prototypes for such conjunctions can be composed from prototypes for their constituents. ... We are concerned only with those composite concepts that do have prototypes, and we have argued elsewhere that such prototypes do not exhaust the contents of a concept (Smith et al., 1988, p. 486).

We thought that this would suffice to communicate the fact that we do not identify concepts with their prototypes. However, in Fodor and Lepore (1996) we read the following:

> There is a standard objection to the idea that concepts might be prototypes (or exemplars, or stereotypes): because they are productive, concepts must be compositional. Prototypes aren't compositional, so concepts can't be prototypes. However, two recent papers (Smith and Osherson, 1984 (OS2); Kamp and Partee, 1995) reconsider this consensus. (pp. 253–254).
>
> But if prototypes aren't compositional, then, to put it mildly, the identification of concepts with prototypes can't explain why concepts are productive. Both OS2 and KP offer solutions for this problem, but it seems to us that neither is even close to satisfactory. (p. 263).

Having misunderstood the purpose of our model, Fodor and Lepore go on to express dismay over its inability to account for conceptual combination and membership –

[4]Alternative perspectives on prototype composition are advanced in Medin and Shoben (1988), Thagard (1984), Murphy (1988), Huttenlocher and Hedges (1994). In Lakoff (1987), concepts are discussed in terms of 'idealized cognitive models', or ICMs. On page 147 we find the following idea. 'In this case the relevant aspects of the evoked ICMs in the striped apple example are our idealized image of stripes and our idealized image of an apple. The Parsimony Principle yields a simple image overlap – an apple with stripes – for our new complex ICM. This is Osherson and Smith's prototypical striped apple, and it works just as it should'. Although there is much of value in Lakoff's discussion, the foregoing theory of complex categories is open to obvious complaints. Thus, nothing is said about the orientation of the stripes relative to the apple, or about their size and placement. For example, when the two images are overlapped, does the apple fall entirely inside a single black stripe? Obviously, Lakoff wishes to place the stripes so as to create the best possible example of striped apple. But this leaves us where we started. Anyway, even after the perfect striped apple lays before the mind's eye, how would it be used to judge relative typicality in the category striped apple? We find no answer in Lakoff's book.

which we have always taken to be unconnected to tvpicalitv (e.g., Osherson and Smith, 1981; Osherson and Smith, 1982).[5]

4.2. *The form of a compositional principle*

Recall that the 'min rule' of fuzzy logic (Zadeh, 1965) can be treated as a hypothesis about typicality by postulating the equation

(6) $c^p_{A\wedge B}(o)=\min[c^p_A(o),\ c^p_B(o)]$, for given adjective-noun concept $A \wedge B$ and object o.

It was observed in Osherson and Smith (1981) that (6) gives the wrong answer about adjective noun concepts like *striped apple* and *pet fish*. For example, goldfish seem to be better examples of the latter concept than they are of either *pet* or *fish*, in violation of (6). This case supports a generalization that will be formulated in Section 5. At present we note only that (6) attempts to derive the value of $c^p_{A\wedge B}(o)$ on the basis of no more than $c^p_A(o)$ and $c^p_B(o)$. So part of the difficulty with (6) might be its restricted form.

K&P seem to have the same idea in mind when they write that 'the supervaluation account is not forced into claiming that the degree to which an object a satisfies a conjunctive concept (A&B) is fully determined by the degrees to which it satisfies A and B' (p. 155). They go on to say (p. 180):

> Indeed, one of the points of this paper has been to show that part of the problem of finding a compositional account of the concept combinations O&S consider is of this very sort. Even if one's primary concern is that of accounting for the c-functions of complex concepts, there is no a priori reason why it should be possible, as O&S seem to implicitly assume, to do this just by looking at the c-functions of the parts.

Let it be noted that the very same point was already raised in Osherson and Smith (1982, pp. 304–306). We there formulated the Simple Functional Hypothesis, namely

> (*SFH*): There is some function f such that $c^p_{A\wedge B}(o)=f[c^p_A(o), c^p_B(o)]$, for given adjective-noun category $A \wedge B$ and object o.

[5]Fodor and Lepore (1996) correctly observe (p. 264) that our model is not adapted to adjective-noun combinations in which the adjective is non-intersective, or in which the adjective denotes properties whose interpretation depends on the noun. However, their remarks do not go beyond the discussion of exactly these points in Smith et al. (1988, pp. 495, 523–525). For example, in the latter article, we discuss cases such as *wooden spoon* – raised previously in Medin and Shoben (1988) – in which the explicit modifier *wooden* interacts with unmentioned properties of spoons such as their size (thus, good examples of *wooden spoon* are not only more wooden than good examples of *spoon*, they are also larger). No additional insight seems to issue from Fodor and Lepore's remarks about the parallel case *male nurse*. For more on the context dependence of adjectives, see Halff et al. (1976), Murphy (1988), Medin and Shoben (1988), Rips (1989); the matter is also discussed in K&P, pp. 142–143. Limitations of our model of prototype combination have led some investigators to underline the role of lay-theories in typicality judgment; see, for example, Medin and Shoben (1988), Murphy (1988), Murphy and Medin (1985).

It is (*SFH*) that K&P attribute to us. In point of fact, Osherson and Smith (1982) presented an argument 'to show generally that the Simple Functional Hypothesis is false, i.e., that no function of the kind specified in (*SFH*) exists' (Osherson and Smith, 1982, p. 305). It was for this explicitly stated reason that the compositional theory advanced in Smith and Osherson (1984) and Smith et al. (1988) does not conform to (*SFH*); indeed, the typicality structure of complex concepts like A $\wedge$ B is there claimed to depend on facts not predicted by c^p_A and c^p_B.[6]

K&P's recalibration proposal has a different character from ours, but it too escapes the narrow confines of (*SFH*). To evaluate their idea, we first indicate the critical data it is designed to explain.

5. The conjunction effect

In Section 6 we argue against thesis (4c). The argument will turn on the following example, introduced in Osherson and Smith (1981). Let a be an irregularly shaped apple with regular stripes. Then it is intuitively clear that

(7) $c^p_{striped\ apple}(a) > c^p_{apple}(a)$.

The inequality (7) illustrates a tendency that might be formulated as follows.

(8) 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.

The phenomenon described in (8) is sometimes called the 'conjunction effect', and has been documented in many studies.[7] Of course, the effect cannot be expected to arise for non-intersective adjectives like 'brilliant', for, John can be a good example of a brilliant person yet a poor ice-skater, and he will not be a good example of *brilliant ice-skater* (at least, on one reading), contrary to (8). The importance of distinguishing intersective adjectives from the rest has been a constant theme in the literature on the typicality of complex concepts.[8]

It appears clearly on pp. 157–158 that an important desideratum for K&P's recalibration theory is to predict (7). They claim to have achieved this by recalibrating the modifier in examples like *striped apple* and then submitting the recalibrated expression to supervaluation theory. They say (p. 166):

> We hope, however, to have indicated how the process of recalibration can be reconstructed in formal terms compatible with the supervaluation method, and

[6]The argument against (SFH) presented in Osherson and Smith (1982) uses the adjective 'round'. Some people have expressed doubts to us that 'round' is intersective, so we observe that the same argument can be reconstructed using a wide variety of concepts – for example, the adjective 'acidic' applied to wine and vinegar. Interpreted as 'having a sour taste' (as specified in the American Heritage Dictionary), 'acidic' is plainly intersective.

how within such an extended formal setting the problem presented by combinations such as *striped apple* finds its resolution.

Let us now see whether recalibration achieves its aim.

6. K&P's recalibration principle: Thesis (4c)

On pages 164–166, K&P consider adjective-noun concepts like *striped apple*, and offer a compositional mechanism for deriving $c^{p}_{striped\ apple}(o)$ from facts about $c^{p}_{striped}$ and c^{p}_{apple}. Their goal is to explain the inequality (7), namely, the greater typicality of our striped apple a with respect to *striped apple* versus *apple*. This is supposed to be achieved via recalibration of $c^{p}_{striped}$ relative to the class of apples (which is assumed to have sharply defined members). With respect to *striped apple*, K&P's equations on p. 165 come to this:

(9) Suppose that x is a genuine apple. Then

$$c^{p}_{striped\ apple}(x) = \frac{c^{p}_{striped}(x) - \min}{\max - \min}$$

where

$\max = \mathit{supremum}\{c^{p}_{striped}(y) \mid y \text{ is a genuine apple}\}$

and

$\min = \mathit{infimum}\{c^{p}_{striped}(y) \mid y \text{ is a genuine apple}\}.$

Intuitively, *max* is the greatest stripedness that apples assume, and *min* is the least. Zadeh advanced a similar scheme in Zadeh (1978, 1982), using the term 'normalization' rather than 'recalibration'. We argued in Osherson and Smith (1982, Section 4.1) that Zadeh's proposal has counterintuitive consequences. It seems to us that (9) fares no better, and for much the same reasons.

Let b be a wretched (but genuine) apple with nice stripes, and let c be a better

[7]See Smith and Osherson (1984), Medin and Shoben (1988), Hampton (1988), Shafir et al. (1990). It is noteworthy that the conjunction effect is also explicitly acknowledged in Lakoff (1987, p. 141). On the other hand, K&P express reservations (p. 170) about the use of line drawings and cardboard cutouts (instead of more natural stimuli) in some psychological experiments that support the conjunction effect. In response, it should be noted that purely verbal stimuli are usually employed. Moreover, one of us recently painted stripes on a warped Golden Delicious apple following the indications left by K&P: 'stripes or streaks virtually always follow longitudinal lines, are most noticeable near the stem end, and are irregular and shaded'. Eleven out of twelve respondants affirmed that the result was a better example of the concept *striped apple* than of *apple*. The twelfth subject expressed uncertainty (and a thirteenth claimed to be unable to make sense of the question).

[8]See, for example, Osherson and Smith (1981, p. 43), Zadeh (1982, p. 292), [Osherson and Smith (1982, p. 301–302), Smith et al. (1988, p. 523–524), Murphy (1988, pp. 535–538), Murphy and Medin (1985, p. 306), Hampton (1988, p. 14). K&P provide further discussion of the same point (pp. 136–145).

apple with the same stripes. Since the stripes are the same, $c^P_{striped}(c) = c^P_{striped}(b)$, so (9) implies

(10) $c^P_{striped\ apple}(c) = c^P_{striped\ apple}(b)$.

But observe that c is a better example of *striped apple* than is b, that is, $c^P_{striped\ apple}(c) > c^P_{striped\ apple}(b)$, contrary to (10). (We hope that our typicality claim needs no argument; it rests on the fact that b and c have the same stripes, but c is a better apple-example than b.)

Moreover, since K&P take the range of c^P functions to be [0,1], the value of *max* must be 1 or very nearly so. For, there exists an apple with nearly perfect stripes. (We are confident of this fact because we drew the stripes ourselves.) Any doubt about the matter can be resolved by purchasing an apple at the grocery store and applying whatever stripes one likes; the thing remains an apple. Similarly, the value of *min* would seem to be 0 since many apples are perfectly unstriped. It thus follows from (9) that

$$c^P_{striped\ apple}(x) = c^P_{striped}(x) \text{ for any genuine apple } x.$$

This seems wrong. For, let a be an apple with imperfect but recognizable stripes. Then it seems clear that $c^P_{striped\ apple}(a) > c^P_{striped}(a)$, in other words, that a is a better example of a striped apple than of a striped thing.

If the reader is uncomfortable with drawing artificial stripes on natural apples, we may change the example to *red apple*. Here there can be no doubt that

$$\max = \mathit{supremum}\{c^P_{red}(y) \mid y \text{ is a genuine apple}\} \approx 1$$

and

$$\min = \mathit{infimum}\{c^P_{red}(y) \mid y \text{ is a genuine apple}\} \approx 0.$$

Indeed, Macintoshes have been bred to be as red as red can be, whereas Granny Smiths are perfectly green. So by K&P's proposal,

(11) $c^P_{red\ apple}(x) = c^P_{red}(x)$ for any genuine apple x.

Now let d be a red but wormy apple, and let e be a beautiful apple of the same color. Then (11), along with the identical colors of the apples, yields $c^P_{red\ apple}(e) = c^P_{red}(e) = c^P_{red}(d) = c^P_{red\ apple}(d)$. However this seems not to respect clear intuitions about exemplification. Among other things, we seem obliged to acknowledge that $c^P_{red\ apple}(e) > c^P_{red\ apple}(d)$, that is, e is a better example of red apple than is d.

Finally, consider a perfect apple q that is red. Since K&P bound their prototype scale at 1, $c^P_{apple}(q) = 1.0$. According to (9) it is thus impossible that $c^P_{red\ apple}(q) > c^P_{apple}(q)$. Yet such is precisely the experimental result obtained in Smith and Osherson (1984, Tables 3 and 5).

In summary, we do not believe that the recalibration scheme embodied in (9) can account for the compositionality of typicality.

7. Typicality versus membership in complex concepts: Thesis (4b)

One reason for skepticism about a robust relation between typicality and vagueness is the conjunction effect, formulated as (8) in Section 5. This phenomenon seems to require that typicality, in contrast to degree of membership, be measured on an unbounded scale. To see this, suppose that the typicality of the prototypical chair x reached the maximum value, say, 1. Since x is a particular object it has a particular color, say, maroon. If the conjunction effect holds in the present case then

(12) $c^p_{maroon\ chair}(x) > c^p_{chair}(x)$.

That is, x is an even better example of *maroon chair* than of *chair*. We believe (12) both on intuitive grounds and because it conforms with the experimental results in Smith and Osherson (1984) for structurally similar cases of the conjunction effect. Of course, (12) shows that $c^p_{chair}(x)$ does not reach a maximum after all. Nor does it seem plausible to affirm that $c^p_{maroon\ chair}(x)$ is maximal, instead of $c^p_{chair}(x)$. This is because the foregoing argument may be repeated at the next level. Since x is a particular object, it is upholstered in a particular material, say leather; and x appears to be an even better example of *maroon leather chair* than of *maroon chair*. We note that the model proposed in Smith and Osherson (1984) and Smith et al. (1988) uses an unbounded scale to measure typicality. This is achieved in a natural way by use of the Contrast Model (Tversky, 1977) to quantify similarity, which itself uses an unbounded scale.[9]

When it comes to membership, in contrast to typicality, x does not seem to belong more to *maroon chair* than to *chair*. It simply belongs unequivocally to both, just as it belongs unequivocally to *maroon leather chair*. So by principle (2), all of $c^e_{chair}(x)$, $c^e_{maroon\ chair}(x)$ and $c^e_{maroon\ leather\ chair}(x)$ reach the same maximal value, conventionally set to 1. This contrast between typicality and membership also applies to *striped apple*, and indeed to any concept *AN* expressed by an intersective adjective *A* and a noun *N* with instances of varying typicality. For any such concept, typicality and membership appear to be scaled differently. So we doubt that $c^p_{AN} = c^e_{AN}$, and hence doubt Thesis (4b).

There is independent reason to be skeptical of such equalities as $c^e_{maroon\ chair} = c^p_{maroon\ chair}$ and $c^e_{chair} = c^p_{chair}$ (the latter is Thesis (4a) applied to *chair*). In the presence of (12) they imply

(13) $c^e_{maroon\ chair}(x) > c^e_{chair}(x)$.

To see that (13) makes a strange claim about membership, notice that the concept *maroon chair* imposes stronger criteria for membership than does the laxer concept *chair*. So it would seem odd that an object could have better credentials in the former extension than in the latter.

[9] That typicality cannot be scaled in a closed interval was observed in our criticism (Osherson and Smith, 1982, pp. 315–316) of G.V. Jones' 'Stack Theory' of prototype combination (Jones, 1982). The use of the Contrast Model to measure typicality was suggested in Tversky (1977, pp. 347–349).

We are led in this way to the provisional conclusion that trying to explain both typicality and vagueness via one, overarching formalism is not the best way to understand either. Although both phenomena admit of graded cases, the gradations seem to bear on distinct facets of human language use and conceptual organization. Moreover, vagueness strikes us as a more fundamental problem for the semantics of natural language than does goodness of exemplification.

The formalism of supervaluations was introduced in van Fraassen (1966), and proposed in Kamp (1975) as an account of vagueness. It is given particularly clear presentation by K&P. So we now propose to consider supervaluation theory on its own merits, divorced from ancillary claims about prototypes. The theory has two potential contributions to make. On the one hand, it can be used to close needless truth gaps. On the other hand, it can be enriched by further assumptions and offered as a model of vagueness. In the next section we examine supervaluations from both perspectives.

8. Supervaluation theory

8.1. *Treatment of logical truth*

We may not know whether a proposition φ is true or false but realize that whichever the case, $\varphi \vee \neg\varphi$ is true and $\varphi \wedge \neg\varphi$ is false. Indeed, we may not even feel that φ has a truth-value (without further clarification), yet nonetheless want to accept $\varphi \vee \neg\varphi$ and deny $\varphi \wedge \neg\varphi$. One virtue of supervaluations is to make this judgment precise and general. If for no other reason, supervaluations greatly illuminate proper reasoning in the presence of missing truth-values.[10]

The latter opinion, however, presupposes that the truth of $\varphi \vee \neg\varphi$ and the falsity of $\varphi \wedge \neg\varphi$ are absolute, admitting no degrees. Such has been our stance from the outset. Thus, instead of perceiving a glimmer of truth in the literal reading of 'Man is both an ape and not an ape', we suggested (Osherson and Smith, 1982, p. 313) that its evident falsehood leads to non-literal interpretation in the manner described in Grice (1975). K&P appear to accept this account of contradictions (pp. 155–156), but there are also many dissenters, notably, the fuzzy logicians.

It is useful to note how committed fuzzy logic is to the non-standard reading of contradictions. Suppose that fuzzy truth-assignment μ over a propositional language has the well-known properties:

(14) (a) $\mu(\theta \wedge \varphi) = \min[\mu(\theta), \mu(\varphi)]$.
(b) $\mu(\neg\theta) = 1 - \mu(\theta)$.

[10] In their logic text, Lambert and van Fraassen put forth the method of supervaluations 'as a device allowing one to tolerate truthvalueless statements while yet holding to the logic developed in this book' (Lambert and van Fraassen, 1972, p. 222). The absence of truth-values at issue for these authors arises from presupposition failure. But supervaluations work just as well if truth-value gaps arise from vagueness.

Then it is easy to prove the following:

> Suppose that $\mu(\theta \wedge \neg\theta) = \mu(\varphi \wedge \neg\varphi)$, for every pair of propositions θ,φ. Then the range of μ is limited to no more than two values.

In other words, in the presence of (14), contradictions must admit a range of truth-values on pain of μ collapsing to a non-fuzzy, two-value truth-assignment. (A similar argument is presented in Elkan (1993).) Moreover, abandoning (14) is a substantial step for fuzzy logic, since (14a) and to a lesser extent (14b) are motivated by compositional principles that fuzzy theorists appear to embrace. (For discussion of the axiomatic foundations of fuzzy operators, see Klir and Folger, 1988, Ch. 2).) All of this is perfectly well understood by the fuzzy logic community, which simply denies that contradictions are literally and identically false (for a trenchant statement, see Dubois and Prade, (1994)).[11]

8.2. *Treatment of vagueness*

In contrast to its perspicuous account of truth-value gaps, K&P's enriched version of supervaluation theory seems to us less successful as an account of vagueness. One difficulty is the requirement that partial models establish sharp positive and negative extensions for predicates. Such sharpness conjures up the same kind of liminal cases that confound traditional semantics. In the domain of men, for example, which are clearly tall? Must they be at least 5 feet 11 inches, or will 5 feet 10.99 inches do as well? This objection to the supervaluational account of vagueness receives the following expression in Timothy Williamson's penetrating study of the sorites paradox of the heap and other continua.

> If it is hopeless to look for the first red shade in a sorites series from orange to red, it is equally hopeless to look for the first shade which can truly be called 'red' (try). The idea that our rough-and-ready use of vague terms does not determine hidden boundaries tells just as much against a pair of hidden second-order boundaries between the true and the neither true nor false and between the latter and the false as it does against one hidden first-order boundary. (Williamson, 1994, p. 157)

K&P seem to acknowledge the difficulty in their footnote 14 (p. 151), and suggest that it can be handled by allowing context to determine the exact range of indeterminacy. This remark leaves us wondering why context can perform the latter task

[11] Lakoff (1987, p. 141) denies the claim in Osherson and Smith (1981) that *apple that is not an apple* is empty perforce. He says: 'Such intuitions have been disputed: a carved wooden apple might be considered an apple that is not an apple ... Osherson and Smith do not consider such possibilities'. The latter remark is accurate. It would never occur to us to classify a carved wooden object of any shape as an apple (any more than we would be tempted to engage conversation with a bust of Aristotle). Such disparity in intuition suggests that participants in the debate about graded membership are talking past each other, referring to different phenomena.

but is unable to finish the job of resolving the remaining vagueness, thereby obviating the need for supervaluations altogether.

Let us nonetheless accept the idea of sharp positive and negative extensions, in order to examine the manner in which supervaluation theory quantifies vagueness. According to the proposal described on pp. 153–155, the value of $c^e_{con}(o)$ in a partial model $\mathcal{M}$ is the measure of the set of completions of $\mathcal{M}$ that (a) respect the semantic constraints imposed by the meanings of *con* and *o* and possibly other terms, and (b) place *o* in the completed, positive extension of *con*. This ingenious idea has intuitive appeal, but it also makes odd predictions when the same objects are situated in different models. We give one example.

Consider the vague predicate *fit*, in the athletic sense. Fitness involves many dimensions of physical prowess, but it will do no harm to limit attention to just two: speed and strength. We will consider three partial models, denoted $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$. The domains of the models consist of men, each of whom will be conceived as a vector (x,y), where x is the distance in miles that the man can run in 10 min, and y is the weight in pounds that the man can lift above his head. Since no particular trade-off between speed and strength is enforced by the meaning of *fit*, the only semantic rule that seems applicable is the following dominance principle.

(15) Suppose that $x \leq x'$ and $y \leq y'$. Then for every completion $\mathcal{M}'$ of a partial model $\mathcal{M}$, if $(x,y) \in \|fit\|^+_{\mathcal{M}'}$, then $(x',y') \in \|fit\|^+_{\mathcal{M}'}$.

Plotting vectors in the plane, rule (15) can be stated geometrically. If point p is below and to the left of point q, and $p \in \|fit\|^+_{\mathcal{M}'}$, then also $q \in \|fit\|^+_{\mathcal{M}'}$

Let $\mathcal{M}_1$ be the partial model with domain $\{a,b,c,d\}$ shown in Fig. 1. Man *a* thus falls into the negative extension of *fit*, man *d* falls into the positive extension, and men *b*,*c* fall into neither. In view of rule (15) there are just three completions of $\mathcal{M}_1$. Two of them make *c* fit, and one makes *b* fit. Proportion of completions is the natural measure for models with finite domains. Using it, we end up in M_1 with $c^e_{fit}(b) = 1/3$ and $c^e_{fit}(c) = 2/3$. Qualitatively, these numbers seem unobjectionable.

Next consider the partial model $\mathcal{M}_2$ in Fig. 1. The additional men *x*, *y*, *z* also have unclear fitness since they fall strictly between the unclear cases *b*, *c*. Counting completions under the constraint (15), we see that $\mathcal{M}_2$ has changed the degrees of membership of the original men *b*, *c* since in $\mathcal{M}_2$ we have $c^e_{fit}(b) = 1/6$ and $c^e_{fit}(c) = 5/6$. Such lability in degree of membership makes us uncomfortable, but it might be argued that the context provided by *x*, *y*, *z* has somehow changed just how fit *b* and *c* really are. However, now consider partial model $\mathcal{M}_3$ in Fig. 1. In contrast to $\mathcal{M}_2$, none of *x*, *y*, *z* dominates the other, although they still fall between *b* and *c*. Again we count the completions consistent with (15), and for this model we find: $c^e_{fit}(b) = 1/10$ and $c^e_{fit}(c) = 9/10$. That is, x, y, z drive b and c further apart in $\mathcal{M}_3$ than in $\mathcal{M}_2$. We are baffled by how supervaluational theorists can deploy context to explain the greater impact of x, y, *z* in $\mathcal{M}_3$ than in $\mathcal{M}_2$. If the context provided by *x*, *y*, *z* really affects the fitness of *b* and *c* then it might have been expected to cause greater separation in $\mathcal{M}_2$ than in $\mathcal{M}_3$. For, stringing out x, y, z between a, *b* as in $\mathcal{M}_2$ produces an ordinal scale of fitness with *a*, *b* near the extremes. In contrast, the ball

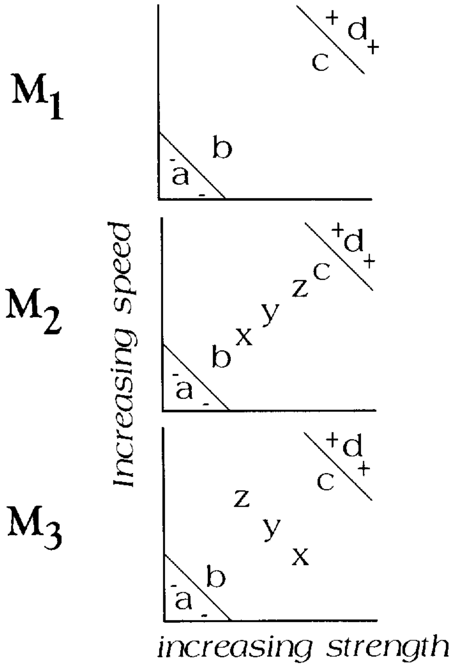

Fig. 1. Three partial models.

of unclear cases shown in $\mathcal{M}_3$ appears to increase overall ambiguity, allowing a, b to be assimilated to each other. As we have seen, greater separation in $\mathcal{M}_2$ than in $\mathcal{M}_3$ is the reverse of what is predicted by the theory.

To avoid such predictions it is tempting to insist that the domains of partial models for *fit* include all possible men. In the present case, this amounts to including every point in the first quadrant of the plane. Degree of fitness may then be computed via an appropriate measure on the class of potential dividing lines that respect (15). Although such an 'intensional' version of supervaluation theory might be feasible for *fit* as conceived above, it seems unworkable for many other predicates. Consider, for example, the vague predicate *intelligent* as applied to men. Application of intensional supervaluation theory here requires that a given partial model include as many possible men as there are distinct things to think about, since degree of intelligence is sensitive to the objects of a person's thought. However, on pain of easily derived paradox, this is too much to place in the domain of any one model.

9. A concluding remark on vagueness and truth

The transition from graded membership to graded truth seems natural and inevitable. It occurs throughout the literature on fuzzy sets (e.g., Dubois and Prade, 1988, p. 292; Klir and Folger, 1988, p. 31) but appears suspect to many others, including us (Osherson and Smith, 1981, Section 3.5) and K&P (pp. 150–151). One source of misgivings can be explained via an example given in Klir and Folger, 1988, p. 32), concerning a 25-year-old woman named 'Tina'. After sketching a smooth curve relating age to membership in *young*, these authors affirm: 'The truth value of 'Tina is young' is 0.87'. The question immediately arises why the youth curve was drawn in such a way as to deny the statement truth value 0.9 or 0.4 in place of the favored 0.87. We have never heard a satisfactory answer to this kind of question, but are informed instead of specific engineering applications of fuzzy logic, including the celebrated fuzzy washing machine.[12]

Note that a probabilistic interpretation of partial truth is not to be recommended, since it would rest on a confusion between uncertainty and indefiniteness. The distinction between the latter concepts is fully appreciated by fuzzy logicians (e.g., Dubois and Prade, 1994, p. 150; Klir and Folger, 1988, Section 5.1).[13]

Refusing to admit graded truth does not, of course, dissolve the perplexity that surrounds vagueness. However, conceiving the latter as connected to category membership rather than to truth might help in evaluating current approaches to vagueness (e.g., Williamson, 1994). It also contributes to the safeguard of classical logic, with its limpid and sensible rules for deriving the truth-values of many compound statements from the truth-values of their parts.

Acknowledgements

We thank Gennaro Chierchia, Jerry Fodor, Susan Gelman, Ernest Lepore, Massimo Piatelli-Palmarini, and Timothy Williamson for helpful discussions.

References

Armstrong, S.L., Gleitman, L.R., Gleitman, H., 1983. What some concepts might not be. Cognition, 13, 263–308.

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

Chierchia, G., McConnell-Ginet, S. 1992. Meaning and Grammar: An Introduction to Semantics. MIT Press, Cambridge, MA.

Dubois, D., Prade, H. 1988. An introduction to possibilistic and fuzzy logics. In: Smets, P., et al. (Eds.), Non Standard Logics for Automated Reasoning. Academic Press, San Diego, CA.

Dubois, D., Prade, H. 1994. Can we enforce full compositionality in uncertainty calculi? In: Proc. 12th Natl. Conf. on Artificial Intelligence, Vol. 1. AAAI Press/MIT Press, Menlo Park, CA, pp. 149–154.

[12]Eschewing fuzzy truth does not entail embracing a definitional theory of meaning. In Hampton (1995, p. 688) the definitional theory is attributed to us on the basis of Osherson and Smith (1981),Osherson and Smith (1982). However, the 'traditional theory' alluded to in the latter papers was intended only in the binary sense, not in the sense of pairing concepts with non-trivial definitions.

[13]Much of the fuzzy logic literature is indeed concerned with uncertainty, in the form of belief and plausibility measures. The Dempster-Shafer theory (Shafer, 1976, Shafer, 1986) is usually taken as the starting point. However, these formalisms are understood to bear on evidence and conviction, not truth (see Smets, 1988).

Elkan, C. 1993. The paradoxical success of fuzzy logic. In: Proc. AAAI'93, Washington, DC, pp. 698–703.
Fodor, J.A., 1981. The present status of the innateness controversy. Representations, ??, 257–316.
Fodor, J., Lepore, E., 1996. The red herring and the pet fish: why concepts still can't be prototypes. Cognition, 58, 253–270.
Grice, H.P. 1975. Logic and conversation. In: Cole, P. and Morgan, J.L. (Eds.), Syntax and Semantics, Vol. 3. Academic Press, New York, pp. 41–58.
Halff, H.M., Ortony, A., Anderson, R.C., 1976. A context-sensitive representation of word-meanings. Memory and Cognition, 4, 378–383.
Hampton, J.A., 1988. Overextension of conjunctive concepts: evidence for a unitary model of concept typicality and class inclusion. Journal of Experimental Psychology: Learning Memory, and Cognition, 14 (1), 12–32.
Hampton, J.A. 1993. Prototype models of concept representation. In: van Mechelen, I., Hampton, J.A., Michalski, R.S., Theuns, P. (Eds.), Categories and Concepts: Theoretical Views and Inductive Data Analysis. Academic Press, London.
Hampton, J.A., 1995. Testing the prototype theory of concepts. Journal of Memory and Language, 34, 686–708.
Huttenlocher, J., Hedges, L., 1994. Combining graded categories. Psychological Review, 101 (1), 157–165.
Jones, G.V., 1982. Stacks not fuzzy sets: an ordinal basis for prototype theory of concepts. Cognition, 12, 281–290.
Kalish, C.W., 1995. Essentialism and graded membership in animal and artifact categories. Memory and Cognition, 23 (3), 335–353.
Kamp, H., Partee, B., 1995. Prototype theory and compositionality. Cognition, 57, 129–191.
Kamp, H. 1975. Two theories about adjectives. In E.L. Keenan (Ed.), Formal Semantics of Natural Languages. Cambridge University Press, Cambridge.
Kay, P., McDaniel, C., 1978. The linguistic significance of the meanings of basic color terms. Language, 54 (3), 610–646.
Klir, G., Folger, T. 1988. Fuzzy Sets, Uncertainty, and Information. Prentice Hall, Englewood Cliffs, NJ.
Lakoff, G. 1987. Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. University of Chicago Press, Chicago, IL.
Lambert, K., van Fraassen, B.C. 1972. Derivation and Counterexample: An Introduction to Philosophical Logic. Dickenson, Encino, CA.
Medin, D., Shoben, E., 1988. Context and structure in conceptual combination. Cognitive Psychology, 20, 158–190.
Murphy, G.L., Medin, D., 1985. The role of theories in conceptual coherence. Psychological Review, 92 (3), 289–316.
Murphy, G.L., 1988. Comprehending complex concepts. Cognitive Science, 12, 529–562.
Osherson, D., Smith, E.E., 1981. On the adequacy of prototype theory as a theory of concepts. Cognition, 9, 35–58.
Osherson, D., Smith, E.E., 1982. Gradedness and conceptual combination. Cognition, 12, 299–318.
Rips, L. 1989. Similarity, typicality, and categorization. In: Vosniadou, S., Ortony, A. (Eds.), Similarity and Analogical Reasoning. Cambridge University Press, Cambridge, UK.
Rosch, E., Mervis, C., 1975. Family resemblance: studies in the internal structure of categories. Cognitive Psychology, 7, 573–605.
Rosch, E. 1973. On the internal structure of perceptual and semantic categories. In: Moore, T.E. (Ed.), Cognitive Development and the Acquisition of Language. Academic Press, New York.
Rosch, E. 1978. Principles of categorization. In Rosch, E., Lloyd, B.B. (Eds.), Cognition and Categorization. Erlbaum, Potomac, MD.
Shafer, G. 1976. A Mathematical Theory of Evidence. Princeton University Press, Princeton, NJ.
Shafer, G., 1986. Constructive probability. Synthese, 48, 1–60.
Shafir, E., Smith, E.E., Osherson, D., 1990. Typicality and reasoning fallacies. Memory and Cognition, 18 (3), 229–239.

Smets, P. 1988. Belief functions. In: Smets, P., Mamdani, A., Dubois, D., Prade, H. (Eds.), Non-Standard Logics for Automated Reasoning. Academic Press, London.

Smith, E.E., Medin, D. 1981. Categories and Concepts. Harvard University Press, Cambridge, MA.

Smith, E.E., Osherson, D., 1984. Conceptual combination with prototype concepts. Cognitive Science, ??, 337–361.

Smith, E.E., Osherson, D. 1988. Compositionality and typicality. In: Schiffer, S., Stelle, S. (Eds.), Cognition and Representation. Westview Press, Boulder, CO.

Smith, E.E., Osherson, D., Rips, L., Keane, M., 1988. Combining prototypes: a selective modification model. Cognitive Science, 12, 485–527.

Thagard, P. 1984. Conceptual combination and scientific discovery. In: Asquith, P., Kitcher, P. (Eds.), PSA, 1. Philosophy of Science Association, East Lansing, MI.

Tversky, A., 1977. Features of similarity. Psychological Review, 84 (4), 327–362.

van Fraassen, B., 1966. Singular terms, truth-value gaps, and free logics. Journal of Philosophy, 63, 481–495.

Williamson, T. 1994. Vagueness. Routledge, London.

Zadeh, L., 1965. Fuzzy sets. Information and Control, 8, 338–353.

Zadeh, L.A., 1978. Pruf – a meaning representation language for natural languages. International Journal of Man-Machine Studies, 10, 395–460.

Zadeh, L.A., 1982. A note on prototype theory and fuzzy sets. Cognition, 12, 291–298.