Concept Membership vs Typicality in Sentence Verification Tasks

Francesca Zarl (francesca.zarl@phd.units.it)
Psychology Unit "Gaetano Kanizsa", Department of Life Sciences,
University of Trieste, via Weiss 21, I - 34128 Trieste, Italy

Danilo Fum (fum@units.it)
Department of Humanistic Studies, University of Trieste
via Tigor 22, I - 34134 Trieste, Italy

Abstract

In the paper we discuss the relation between fuzzy sets and the graded membership and typicality effects found in the study of concepts. After a short overview of the topic, we present three experiments, carried out using the same method but with different situational contexts, which examine whether graded membership and typicality could be considered as independent factors capable of influencing the performance of human participants involved in sentence verification tasks, or they are somehow interrelated. The paper concludes with a general discussion of the experimental findings and the problems they pose for models of concepts based on the theory fuzzy sets.

Keywords: Concept representation; fuzzy set theory; graded membership; vagueness; typicality; sentence verification, categorization.

Introduction

The study of concepts has constituted a key point for psychological research since its very beginning. In the last decade or so, it has acquired new importance in the light of the Semantic Web endeavor to find a computationally effective way to model the ontologies machines need to perform satisfactorily at cognitive-related tasks.

The so-called classical view (Smith & Medin, 1981) regards concepts as rules or definitions, i.e., as sets of properties which are individually necessary and jointly sufficient to determine whether or not a given entity is an instance of the concept in question. From a formal point of view, concepts can be modeled using sets whose intensions are defined by the rules. According to the law of excluded middle, an element is a member of a set (if it meets all the necessary properties) or it is not (if there is at least one necessary property which is not met). Moreover, since all the elements of a set satisfy the same conditions, there is no difference among them as far as their membership is concerned.

The classical view sounds reasonable and intuitive but it is in fact inadequate to explain several empirical phenomena psychologists have found by running their experiments. A number of empirical studies (e.g., Barsalou, 1989) showed that people, required to define a concept, have difficulty in generating lists of properties that are necessary and sufficient for it. Morevover, in trying to identify the features characterizing many everyday concepts, people disagree with each other, and sometimes they disagree also with themselves, with the same person generating different lists on diverse occasions (Bellezza, 1984).

These results depend on the fact that most of our concepts are vague, and they do not have clearly defined boundaries. If vagueness constitutes sometimes a definite asset, it has the disadvantage that none of our categories will ever fit completely with the world, and there will be always cases in which it will be difficult to discriminate whether an instance belongs to a concept or not.

Hampton (1993) found, for example, that people consider some entities as just barely members of a category and other entities as just merely non-members. Members and non-members form a continuum without a clear distinction between them. While it is easy to classify unambiguous elements, people face serious difficulties in determining whether an in-between element does belong or not to a particular category. McCloskey and Glucksberg (1978) found that people, asked to make membership decisions for exemplar-category pairs in two separate sessions, agreed with each other and were consistent between sessions for items clearly related or unrelated with the category (e.g., *apple-fruit* and *cucumber-furniture*, respectively), while they disagreed and were frequently inconsistent in the case of borderline elements (e.g, *curtains-furniture*).

Another embarrassing result for the classical view is constituted by the fact that people consider certain entities as better exemplars of a concept than others; for instance, *sparrows* and *robins* are somehow considered as better *birds* than *ostrichs* or *penguins* (Rosch, 1973). This finding, known as the typicality effect, represents one of the most common and most robust phenomena found in the study of concepts.

It has been found that typicality affects people's performance in a variety of ways and in wide range of cognitive tasks. Typical exemplars of a category are classified faster and more accurately than atypical ones (Rips, Shoben, & Smith, 1973); they are more likely to be considered as members of the category (Hampton, 1979); they are mentioned more frequently when asked to name the members of the category (Mervis, Catlin, & Rosch, 1976); they are the first to be learned in artificial category learning tasks (Rosch, Simpson, & Miller, 1976); they support better analogical inferences (Rips, 1975), etc.

Being incompatible with the classical view of concepts, graded membership and typicality cannot be modeled according to the set-theoretic account on which such a theory is based. Soon after the discovery of these phenomena, an alternative framework for concepts that relied on the idea of fuzzy sets (Zadeh, 1965) was put forward.

In the paper we discuss the relation between fuzzy sets

and the graded membership and typicality effects found in the study of concepts. After a short overview of the topic, we present three experiments, carried out using the same method but with different situational contexts, which examine whether graded membership and typicality could be considered as independent factors capable of influencing the performance of human participants involved in sentence verification tasks, or they are somehow interrelated. The paper concludes with a general discussion of the experimental findings and the problems they pose for models of concepts based on the theory fuzzy sets.

Graded membership, typicality, and fuzzy sets

Fuzzy sets were introduced by Zadeh (1965) to overcome the limitations of the traditional set-theoretic approach in dealing with classes of objects that are not clearly defined. Because these classes are pervasive in all the processes that somehow involve language, fuzzy sets have been considered since their beginning as a promising formalism to represent conceptual knowledge. Following the publication of a critical paper by Osherson and Smith (1981), the interest for fuzzy sets rapidly declined among cognitive scientists determining their practically disappearance from the literature. Recently, however, a recent reevaluation of the topic by Belohlavek and Klir (2011) showed that many arguments raised against fuzzy set were fallacious, and contributed to restate their relevance for the study of concepts.

Fuzzy sets can be conceived as an extension of classical sets. While the characteristic function of the latter maps its domain into the set $\{0, 1\}$, comprising only two elements, the range of the former's function contains an infinite number of elements: all the real numbers in the interval [0,1], including its limits. In this sense, a classical set is a subset of a fuzzy one. While classical sets consider 0 and 1 symbolically, as Boolean entities, the degrees of membership of fuzzy sets have a true numerical value. One important consequence of this fact is that fuzzy sets can be mathematically manipulated in ways that are not allowed by their classical counterpart.

It is evident that both the graded membership and the typicality effect found in the study of concepts can be modeled through fuzzy sets by introducing a function having as range the interval of real numbers [0,1]. The question to be asked is whether graded membership and typicality should be considered as separate phenomena which however depend on an underlying common factor, or whether they denote distinct dimensions to be captured by different functions.

According to Cai, Yeung, and Leung (2012) graded membership and typicality have different nature and are not necessarily related. An instance of a concept can in fact be assigned a high degree of membership and a low degree of typicality (as it is the case for *ostrichs* as members of the class of *birds*) while, on the other hand, there may be entities that have a membership degree close to zero and non-null typicality. This latter case occurs when people consider some entities as exemplars of a given concept (for example, consider *whales* as *fish* or *tomatoes* as *vegetable*) while, technically, they cannot be considered as such. As stated by the authors, graded membership and typicality play different roles in determining whether an entity is an instance of a concept, and are computed in different ways. To determine the degree of membership, the necessary conditions which define a concept according to the classical view are taken into account. Typicality provides an additional mechanism to rank those instances that meet all the membership requirements (and whose values approximate therefore 1). In this case the non-defining features which are widely shared among the set members are tallied. Sometimes this leads to the assignment of typicality values to entities that, without being members of the set, partake of its characteristic properties.

Kamp and Partee (1995) too deny that a single measure could serve the purposes of quantifying both the degree to which an entity is an instance of a given concept and its typicality, interpreted as the degree of proximity to the best example (or prototype) of the concept. The question they ask is whether and how the two functions could be considered as connected, and in particular for which concepts, if any, do they coincide. As an example of lack of interrelation Kamp and Partee refer to the concept of *male nurse*. In this case, according to the authors, knowing the degree of membership—which depends on the intersection of the classes of *males* and *nurses*—does not help in any way to establish typicality. On the other hand, there are cases (for instance, for the concepts of *red* or *chair*) in which knowing the typicality value of an instance helps to establish its membership degree.

Another interesting point of view on this issue is raised by Hampton (2011) who considers typicality and graded membership as separate functions that are based, however, on the same underlying similarity measure. More particularly, "[t]ypicality is a monotonically rising function of similarity, whereas membership is a nondecreasing function of similarity that starts at 0, starts to rise at a certain point k_1, and then reaches a ceiling of 1 at a further point k_2, where k_1 and k_2 are above the minimum and below the maximum values that similarity can take." (Hampton, 2011, p.219)

To better define the relationship between graded membership and typicality, and their possibile connections with fuzzy sets, we realized the following experiments.

Experiments

All the experiments investigated the role that graded membership and typicality play in determining the behavior of people engaged in sentence verification tasks. The experiments asked participants how much they agreed with a series of sentences claiming that *"...Xs are/are not Ys"*. Beyond asserting or negating the membership of X to Y, the sentences differed in the relationship that connected an instance with its putative category. X could be in fact a typical member of Y, an atypical member, a non-member sharing common features with the members of Y, and a non-member without any evident relation with Y. In each experiment, therefore, member-

ship and typicality were orthogonally varied, as it was varied the polarity (affirmative vs negative) of each sentence.

The experiments differed according to the context in which membership judgments were made. It is reasonable to assume that the degree of agreement with a statement could depend not only on the intrinsic relationship between an entity and a class but also on the particular viewpoint taken by participants, or by the specific circumstances in which the sentences are evaluated. (For a review of studies on concept classification manipulating the context of the task, see: Murphy, 2002, pp. 413-422.)

By adopting different kinds of contexts, we tried to determine the generality and robustness of any influence membership and typicality could have on the concept categorization processes underlying the sentence verification task.

Experiment 1

In Experiment 1 we investigated whether the effect of membership and typicality could be modulated by providing participants with a purposive context for sentence evaluation. Our approach is similar to that employed by Hampton, Dubois, and Yeh (2006) who adopted instructions contrasting a purely pragmatic classification context with a more technical one, and compared them with a no-context condition. Contrary to the authors' expectations, however, none of the dependent measures was influenced by the context. We conjectured that, by emphasizing the context also in the sentence text, we could differently influence the criteria participants used to evaluate the sentences.

Participants. Sixty University of Trieste students (48 females), whose age varied from 18 to 53 years (mean = 23.1; sd = 8.0), participated to the experiment. Participants were randomly assigned to three experimental conditions named *Technical*, *CommonSense*, and *Neutral*, respectively.

Design. A 3x2x2x2 mixed design was adopted having *Context* (*Technical, CommonSense*, and *Neutral*) as a between-subjects factor and *Membership* (*Yes* vs *No*), *Typicality* (*Yes* vs *No*), and *Polarity* (*Positive* vs *Negative*) as factors within.

Materials and procedure. In the experiment 64 Italian sentences – the same used in Zarl and Fum (2014) – each involving a relation between an instance and a category, were used. The sentences were divided into 32 pairs. A sentence in each pair made an affirmative statement while the other negated it. The sentences were constructed by balancing the gender and the number of instances and categories which were both of natural (e.g., *tomato-fruit*) and artificial (e.g., *volleyball-sport*) kind. Eight different types of sentences were constructed by varying the three distinct factors of *Membership, Typicality,* and *Polarity*. Table 1 provides an example for each kind of sentence.

Positive sentences are labeled with P while negative sentences are labeled with N. M means that the instance is a member of the category while while $\overline{\text{M}}$ negates it. Analogously, T means that the instance shares some common features with members of the category while $\overline{\text{T}}$ negates any similarity between the instance and category.

Table 1: A sample of sentences used in the experiment.

Sentence kind	Text
PMT	... *canaries are birds*
PM$\overline{\text{T}}$	... *penguins are birds*
P$\overline{\text{M}}$T	... *bats are birds*
P$\overline{\text{M}}\overline{\text{T}}$	... *toads are birds*
NMT	... *toads are not birds*
NM$\overline{\text{T}}$	... *bats are not birds*
N$\overline{\text{M}}$T	... *penguins are not birds*
N$\overline{\text{M}}\overline{\text{T}}$	... *canaries are not birds*

The interpretation of positive sentences is straighforward. PMT means that the instance is a typical member of the category, PM$\overline{\text{T}}$ that it is an atypical member, etc. The labeling of negative sentences, obtained by negating the labels of the positive ones, is based on the criteria participants would follow in agreeing with the sentence content. So, for example, accepting the N$\overline{\text{M}}$T sentence *"... penguins are not birds"*, which negates the PM$\overline{\text{T}}$ *"... penguins are birds"*, means denying membership while acknowledging typicality as evaluation criterion because *penguins*, even if they are in fact *birds*, lack some features that are typical of this category.

The sentences of each experimental condition were introduced by a different phrase which provided a context for their reading. In the case of the *Technical* group, the sentences began with the expression *"In a technical sense ..."*, while those of the *CommonSense* group were introduced by the words *"According to common sense ..."*. For the *Neutral* condition we borrowed the phrase *"In a sense ..."* by Machery and Seppälä (2011) who used it to allow different interpretations for the concepts of their experiments.

The context according to which participants were asked to evaluate the sentences was varied in the instructions, too. For the *Technical* group, the instructions highlighted the fact that concepts are structured according to a taxonomy based on strict membership rules. Participants assigned to the *CommonSense* condition were said that a looser interpretation of concepts would take into account the similarity that exists between them. Finally, the instructions for the *Neutral* group were quite general and did not provide any specific indication about the setting to be adopted.

All the sentences were gathered in a leaflet whose pages contained eight sentences drawn randomly from the total pool. Next to each sentence was printed a 7 cm line whose extreme points were marked with the labels *Completely disagree* and *Completely agree*, respectively. Participants had to indicate their degree of agreement with the sentence by putting a vertical mark on the line. The position of the mark was measured at the next millimeter and converted into a score in the [0,70] interval of integers.

Results A four-way mixed ANOVA with one variable between-subjects (*Context*) and three variables within-

subjects (*Membership, Typicality, and Polarity)* was performed on the experimental data. All the main effects of the within-subjects factors were significant: Sentences in which the instance was a member of the category obtained higher ratings than sentences in which the instance was a non-member ($F_{(1,57)} = 338.67, p < .001$). Analogously, scores for sentences in which the instance was similar to the category typical members were higher than those of sentences in which there was no similarity between the instance and the category ($F_{(1,57)} = 443.91, p < .001$). Finally, affirmative sentences obtained higher judgments than negative ones ($F_{(1,57)} = 64.10, p < .001$). On the other hand, the different contexts did not have any effect on the participants performance, and did not lead to any significant interaction with the other factors.

Table 2 reports the mean scores for the different sentence kinds. To facilitate the understanding of the data, the contextual conditions have been collapsed in the table.

Table 2: Average scores – contextual conditions collapsed.

	Positive		Negative	
	$\overline{\text{M}}$	M	$\overline{\text{M}}$	M
$\overline{\text{T}}$	13.21	44.27	9.21	33.46
T	36.00	62.80	27.79	54.50

An interesting (and puzzling) three way interaction *Membership x Typicality x Polarity* ($F_{(1,57)} = 4.58, p < .05$) was found among the within-subjects variables. The interaction effect has quite a small magnitude but it constitutes an original and unexpected result that requires an adequate explanation (see: the *General discussion and conclusions* section).

To determine which criterion, between *Membership* and *Typicality*, was more influential in determining the participants judgments, a second analysis of the data was made taking into account only the sentences is which the two criteria were directly contrasted. A two way mixed ANOVA was carried out having *Context* as a variable between, and *Criterion* (i.e., *Membership*, for the PM$\overline{\text{T}}$ and NM$\overline{\text{T}}$ sentences, vs *Typicality*, for the P$\overline{\text{M}}$T and P$\overline{\text{M}}$T ones) as variable within. The ANOVA revealed only the significant main effect ($F_{(1,57)} = 19.13, p < .001$) of *Criterion*, with *Membership* sentences obtaining significantly higher ratings than *Typicality* ones (average scores of 38.86 and 31.90, respectively).

Experiment 2

The experiment aimed to assess whether the influence of *Membership* and *Typicality*, and their possible interaction with *Polarity*, could depend on how the material was presented. In a previous work (Zarl & Fum, 2014), we found that participants were more likely to agree with sentences that contradicted each other (e.g., "*In a sense penguins are birds*" vs "*In a sense penguins are not birds*"), and to assigning therefore high scores to both, when the sentences were submitted separately in random order, while contradiction acceptance was lower when they were presented together for the evaluation. Experiment 2 was therefore a control experiment intended to ascertain the generality of the effects of Experiment 1, and to confirm maybe its findings.

Participants. Forty voluntary participants (31 females) from the Trieste area whose age varied from 20 to 42 years (mean = 26.1, sd = 4.8) were engaged in the experiment. Participants were randomly assigned to two experimental conditions named *Pair* vs *Single*.

Design. A 2x2x2x2 mixed design was adopted with *Context* (*Pair* vs *Single*) as a variable between, and *Membership*, *Typicality* and *Polarity* as variables within subjects.

Materials and Procedure. The materials and the instructions of the *Neutral* condition of Experiment 1 were used, with all the sentences introduced by the phrase "*In a sense* ..." and no particular hint provided for the their interpretation. For participants in the *Single* condition, the procedure too was identical to that of the previous experiment, with the sentences presented in random order. The positive and negative sentences of each pair were instead presented together, one below the other, to participants of the *Pair* condition. For this condition, the pairs presentation order, and the order of presentation of the positive and negative sentence within each pair were determined randomly.

Results. The experiment replicated all the findings of Experiment 1. A four-way ANOVA showed the main effects of *Membership* ($F_{(1,38)} = 181.04, p < .001$), *Typicality* ($F_{(1,38)} = 124.872, p < .001$), and *Polarity* ($F_{(1,38)} = 52.889, p < .001$) only. Table 3 shows the mean scores for the different sentences. Similarly to what was done in Experiment 1, because the presentation *Context* did not proved significant, the data are presented without taking it into account.

Table 3: Average scores – contextual conditions collapsed.

	Positive		Negative	
	$\overline{\text{M}}$	M	$\overline{\text{M}}$	M
$\overline{\text{T}}$	14.49	50.15	7.31	37.38
T	36.12	59.82	25.72	55.26

The ANOVA confirmed also the three-way interaction *Membership* x *Typicality* x *Polarity* ($F_{(1:38)} = 14.21, p < .001$) previously found in Experiment 1 that becomes here more perspicuous: the gain due to *Typicality* is clearly greater for the non-members $\overline{\text{M}}$, in the case of *Positive* sentences, and for members M, when the sentences are *Negative*.

Another interesting interaction *Context* x *Membership* ($F_{(1,38)} = 6.25, p < .05$) found by the ANOVA and absent in Experiment 1 is reported in Table 4.

According to the data, when the sentences are presented simultaneously in pairs, the participants judgments become more polarized (i.e. with higher scores for M sentences and lower scores for the $\overline{\text{M}}$ ones) in comparison with the case in

Table 4: The *Context* x *Membership* interaction

	Member M	Non-Member $\overline{\text{M}}$
Pair	58.88	21.20
Single	51.64	30.25

which the sentences are presented singularly and in a randomized order. This polarization effect obtained in the *Pair* condition confirms what had already been pointed out previously by Zarl and Fum (2014).

Finally, a second ANOVA taking into account the positive and negative sentences of the M$\overline{\text{T}}$ and $\overline{\text{M}}$T kind was performed to determine which criterion, between *Membership* and *Typicality*, was more influential in determining the participants judgments. This mixed two-way ANOVA was carried out having *Context* as between-subjects factor and *Criterion* as factor within. The ANOVA revealed only the significant main effect ($F_{(1,38)} = 31.32, p < .001$) of *Criterion*, with *Membership* sentences (mean = 43.76) obtaining significantly higher ratings than *Typicality* ones (mean = 30.92).

Experiment 3

A final experiment was carried out to check whether the previous findings could somehow depend on the fact that all the sentences were always preceded by an introductory phrase. The idea for this control experiment derived from some occasional remarks made by participants in Experiment 2 and by those assigned to the *Neutral* condition of Experiment 1 who, after reading some sentences introduced by *"In a sense ... "*, asked "In which sense?". Experiment 3 thus contrasted the simple assertion or negation of membership with the more elaborate formulation adopted to induce a neutral context for sentence interpretation.

Participants. The participants were 64 residents in the Friuli Venezia Giulia region in Italy (49 females) whose age varied from 19 to 63 years (mean = 25.74; sd = 8.2).

Design. The same 2x2x2x2 mixed design of Experiment 1 was adopted. In this case the levels of the between-subjects *Context* variable were *Neutral* and *NoContext*, respectively.

Materials and procedure. For participants in the *Neutral* condition, the materials and the procedure were identical to those of the *Single* condition of Experiment 2 (and of the *Neutral* condition of Experiment 1). Participants in the *NoContext* condition read the same sentences with the introductory phrase *"In a sense ... "* removed (e.g. *"Penguins are birds", "Canaries are not birds"* etc.).

Results. The same pattern of results found in the previous experiments was replicated in Experiment 3, too. More particularly, the *Context* factor did not have any statistically significant effect on the participants performance indicating that the previous findings were not influenced by the presence of the introductory phrase *"In a sense ... "*. Table 5 reports the average scores for the different conditions.

Again, a four-way ANOVA revealed the significative main

Table 5: Average scores–contextual conditions collapsed.

	Positive		Negative	
	$\overline{\text{M}}$	M	$\overline{\text{M}}$	M
$\overline{\text{T}}$	11.68	47.62	5.94	32.99
T	37.05	64.05	22.38	54.64

effects of *Membership* ($F_{(1,62)} = 433.67, p < .001$), *Typicality* ($F_{(1,62)} = 585.33, p < .001$), and *Polarity* ($F_{(1,6262)} = 188.44, p < .001$), and their three-way interaction ($F_{(1,62)} = 19.72, p < .001$) indicating the greater advantage due to Typicality that occurs in the *Positive* $\overline{\text{M}}$ sentences.

A mixed two-way ANOVA, using *Context* and *Criterion* as factors was carried out on sentences contrasting directly *Membership* and *Typicality*. It showed a significant ($F_{(1,62)} = 40.48, p < .001$)) main effect of *Criterion*, obtaing the former (mean = 40.31) higher scores than the latter (mean = 29.72).

General discussion and conclusions

In the previous section we have reported three experiments, all leading to concordant results. First of all, the experiments confirmed the graded membership and the typicality effects characterizing concepts, and their importance in determining the degree of agreement with a series of statements asserting or denying the membership of an instance to a class.

Between the two factors, membership played the dominant role. In all the experiments, when the factors were directly contrasted, participants assigned higher scores to sentences in which the instances were atypical members of the class in comparison with those in which the instances simply shared some similarity with the class members.

This is a general result that was not affected by any attempt to modulate it by providing participants with different evaluation contexts. Even when explicitly instructed to take into account the similarity between the instance and the typical exemplars of the class (Experiment 1), participants were guided by the membership criterion. In Experiment 2 the simultaneous comparison between contradictory statements led to a polarization of judgments (with higher scores for sentences asserting membership and lower scores for those denying it, when the instance was indeed a member of the class) but did not change the general pattern of results. The same was true in Experiment 3 when assertions and negations were expressed directly in the sentences or were somehow dampened by the introductory phrase *"In a sense"*.

Membership and typicality are not independent factors but they interact with each other in a subtle way. In case of affirmative sentences, the gain in the score due to typicality is higher for non-members than for the members of a category while it is virtually identical when the sentences are negative. This is the most interesting and original finding deriving from all the experiments and, at least to our knowledge, no existing model of concepts is able to predict and explain it.

The results of our experiments highlight some interesting issues for models of concepts based on the theory of fuzzy sets. Due to space limitations we mention only two, both

related to the interpretation of negative sentences.

From a strictly logical point of view, denying that an instance is a member of a set can be considered as equivalent to stating that it is a member of the set complement: negating that *bats* are *birds* is like claiming that they are *non-birds*. Concepts such as *not a bird*, defined in a negative way, are often cited (Connolly, Fodor, Gleitman, & Gleitman, 2007) as examples of concepts for which no prototype exists. Although it may be relatively easy to establish whether a certain instance is a *non-bird*, in this determination the similarity between the instance and *non-birds* should not play any role. The problem is that in our experiments typicality manifests its effect also in negative sentences: the $\overline{\text{NMT}}$ *"Penguins are not birds"* receives higher scores than the $\overline{\text{NMT}}$, *"Canaries not are birds"*, and the same is true for the sentences whose instance is a member of the category.

It is possibly true, as state by a paper reviewer, that "the fact that typicality is present in negative sentences does not mean that it is the effect of any prototype for the negation of the concept. It just means that participants are still capable of evaluating the distance of a nonmember to the prototype of the positive category in terms of similarity". The fact is that what appears as a strong pragmatic effect of negation cannot be easily adapted into the mathematical theory of fuzzy sets.

A second issue concerns the scores given to a sentence and its negation. The function defining the complement of a fuzzy set is monotonically decreasing, i.e. for every instance a, b of a set A and their respective membership degrees, $\mu(a)$ and $\mu(b)$, if $\mu(a) < \mu(b)$ then $c(\mu(b)) \geq c(\mu(a))$, where c is the function defining the complement of A. While this relation is maintained in the aggregated data, in many cases the individual scores given by participants do not respect this property, a fact that casts some further doubts about the possibility of translating directly the axioms of the mathematical theory of fuzzy sets into predictions about the participants behavior.

References

Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76–121). New York: Cambridge University Press.

Bellezza, F. S. (1984). Reliability of retrieval from semantic memory: Common categories. *Bulletin of the Psychonomic Society*, *22*, 324–326.

Belohlavek, R., & Klir, G. J. (2011). Fallacious perceptions of fuzzy logic in the psychology of concepts. In R. Belohlavek & G. J. Klir (Eds.), *Concepts and fuzzy logi* (pp. 121–148). Cambridge, MA: The MIT Press.

Cai, Y., Yeung, C. M. A., & Leung, H. F. (2012). *Fuzzy computational ontologies in context*. Beijing and Berlin: Higher Education Press and Springer.

Connolly, A., Fodor, J. A., Gleitman, L. R., & Gleitman, H. (2007). Why stereotypes don't even make good defaults. *Cognition*, *103*, 1-22.

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

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

Hampton, J. A. (2011). Conceptual combinations and fuzzy logic. In R. Belohlavek & G. J. Klir (Eds.), *Concepts and fuzzy logic* (pp. 209–231). Cambridge, MA: The MIT Press.

Hampton, J. A., Dubois, D., & Yeh, W. (2006). Effects of classification context on categorization in natural categories. *Memory & Cognition*, *34*, 14319–14436.

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

Machery, E., & Seppälä, S. (2011). Against hybrid theories of concepts. *Anthropology and Philosophy*, *10*, 97–126.

McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy set? *Memory & Cognition*, *6*, 462–472.

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationship among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, *7*, 283–284.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

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

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*, 665–681.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*, 1–20.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. M. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic Press.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 491–502.

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

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

Zarl, F., & Fum, D. (2014). Theories of concepts and contradiction acceptance. In P. Bernardis, C. Fantoni, & W. Gerbino (Eds.), *TSPC2014: Proceedings of the Trieste Symposium on Perception and Cognition* (pp. 157–161). Trieste, Italy: EUT: Edizioni Università di Trieste.