Fuzzy Concept Combination: An Empirical Study

Article *in* Fuzzy Sets and Systems · March 2020

DOI: 10.1016/j.fss.2020.03.004

Igor Douven
French National Centre for Scientific Research

Fuzzy Concept Combination: An Empirical Study*

Igor Douven
SND / CNRS / Sorbonne University
igor.douven@sorbonne-universite.fr

Abstract

Fuzzy set theory has been criticized for building on an ill-understood notion, viz., that of graded membership. To address this concern, recent work has proposed an operational definition of this notion, which experimental studies have subsequently shown to make accurate predictions regarding people's judgments of degrees of membership. It is still an open question whether analogous positive results can be obtained for more complex fuzzy concepts, for instance, concepts that can be modeled as intersections of fuzzy sets. The present paper makes a beginning with answering this question. It reports results of an experimental study that derived degrees of membership for a conjunctive concept from the outcomes of the aforementioned studies and then compared these predicted degrees with the degrees as judged by participants. Degrees of membership were derived using all of the better-known fuzzy intersection operators, and thus the results of the study also shed new light on the question of the empirical adequacy of those operators.

Keywords: conceptual spaces; degree of membership; fuzzy logic; fuzzy set theory; intersection operators.

1 Introduction

Classical set theory and classical logic were devised to facilitate mathematical reasoning. As such, they have proven to be powerful tools. It thus makes sense to try to apply them outside the realm of mathematics. But natural language differs in important respects from the language of mathematics. Consider, for instance, the debate about conditionals. "If" in the language of mathematics may be adequately formalized by the material conditional, but few nowadays hold that this is also true of "if" as we use this word colloquially (Evans & Over 2004; Douven 2016a). To give another example: Pragmatics, while it is crucial to the understanding of everyday language (Levinson 2000), is generally believed to have no bearing on the language of mathematics—the assertion that some integers are positive does *not* implicate (the true proposition) that not all integers are positive. Finally, and most relevantly for present purposes, mathematical predicates have well-delineated, crisp extensions, something that is untrue for many or even most natural-language predicates, which tend to be vague to various degrees. Perhaps more than any other, this last point may limit the applicability of logic in domains beyond mathematics, and may, as Russell (1923, p. 65) put it, make logic "not applicable to this terrestrial life but only to an imagined celestial existence."

The wish to overcome this limitation, and to develop formal tools for the semantic representation of vague predicates or concepts (thought of as the mental correlates of predicates), led Zadeh (1965) to the introduction of fuzzy set theory, whose hallmark is a graded membership relation. Where Ω is a class of objects, a fuzzy set $\tilde{S}$ in Ω is defined as a (classical) set of ordered pairs, $\{\langle\omega, \mu_{\tilde{S}}(\omega)\rangle \mid \omega \in \Omega\}$. Here, $\mu_{\tilde{S}}(\cdot)$ is a function from Ω into the $[0, 1]$ interval, which indicates, for each $\omega \in \Omega$, the degree to which ω belongs to $\tilde{S}$. The intuitive interpretation is that the closer $\mu_{\tilde{S}}(\omega)$ is to 1, the "more" ω belongs to $\tilde{S}$, with the extremes 0 and 1 indicating full non-membership and full membership, respectively.

Zadeh argued that the key notions of standard set theory all have natural extensions in fuzzy set theory. For instance, he defined the complement of a fuzzy set $\tilde{S}$ as $\{\langle\omega, 1 - \mu_{\tilde{S}}(\omega)\rangle \mid \omega \in \Omega\}$, and the intersection $\tilde{S}_1 \cap \tilde{S}_2$ of fuzzy sets $\tilde{S}_1$ and $\tilde{S}_2$ as $\{\langle\omega, \min[\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)]\rangle \mid \omega \in \Omega\}$.

Zadeh's 1965 paper was the beginning of a highly successful research program, now generally known as "fuzzy mathematics." This program led to the development of fuzzy algebra, fuzzy calculus, fuzzy graph theory, fuzzy logic, fuzzy probability theory, and much more besides (Dubois & Prade 1979, 1982a; Bělohlávek, Dauben, & Klir 2017, Ch. 5). The many real-world applications the program turned out to have—in such diverse areas as bio-informatics, consumer electronics, decision theory, stock market forecasting, and traffic control—further contributed to its success (Klir & Folger 1988; Klir & Yuan 1995, Pt. II; Zimmermann 2001, Ch. 15; Bělohlávek, Dauben, & Klir 2017, Ch. 6).

Meanwhile, however, the program has been frowned upon by mathematicians, psychologists, and philosophers in almost equal measure. These critics' main concern has been that the notion of graded membership, which figures as a primitive in fuzzy set theory, is too subjective and, well, too fuzzy to play any serious scientific role. What does it *mean*—the critics ask (e.g., Lindley 2004)—to say that an item belongs to a set to some intermediate degree? Relatedly, Haack (1979, p. 441) complains that "fuzzy logic *compromises* with vagueness; it is not just a logic of vagueness . . . it is a vague logic." To be sure, the membership relation of classical set theory serves as a primitive notion as well, but no one doubts that we have a firm pretheoretical grip on that notion, nor does anyone hold that the accompanying Boolean logic is vague.[1]

Recent theoretical work has sought to make graded membership a scientifically respectable notion by grounding it in the conceptual spaces framework, in which concepts are represented geometrically. Said theoretical work used a combination of geometry and measure theory to define graded membership *operationally*, by specifying a procedure for measuring the degrees to which items fall under a given concept. This procedure can also be interpreted as measuring the degree to which an item belongs to the extension of a concept, thus effectively yielding an operational definition of graded membership.

Operationalizing graded membership can only be a first step, of course. Zadeh emphasized right at the start of his 1965 paper that he intended fuzzy set theory to be applicable to real-world phenomena. We must, therefore, also ask how well the proposed operationalization of graded membership succeeds in modeling those phenomena. If, using that definition, we arrive at degrees of membership that are consistently, or even just frequently, out of step with people's judgments concerning what those degrees are, the definition will not have done much to elevate the status of fuzzy set theory.

In recent empirical work (Douven 2016b; Douven et al. 2017), this operational definition was put to the test. While it passed the tests successfully, this amounts at most to a partial vindication of Zadeh's original proposal. The empirical work focused strictly on atomic concepts, and thus left open the question of the empirical adequacy of the various set-theoretic operators, such as the complement

[1] See Bělohlávek, Dauben, and Klir (2017, Ch. 1) for a thorough historical account of the (often skeptical) reception of fuzzy set theory in various scientific quarters.

and intersection operators stated above, as applied to fuzzy sets. Indeed, it is unknown whether, assuming the same operational definition of graded membership, we can predict how people use "not," "and," etcetera, in the context of fuzzy predicates. For instance, supposing that our measure yields $\mu_{\text{nice}}(\text{Bob}) = x$ and $\mu_{\text{athlete}}(\text{Bob}) = y$, can we safely predict that Bob will be deemed a nice athlete to a degree of $\min(x, y)$, in accordance with Zadeh's definition of fuzzy set intersection?

In fact, since Zadeh's initial proposal, many other intersection operators for fuzzy sets have been proposed (see Sect. 4.1; also Paris 2000 and Fermüller 2015). This paper concentrates on such "fuzzy and operators," leaving an investigation of other operators (e.g., "fuzzy not," "fuzzy or," "fuzzy if") for future work. The main research question is whether, for some known intersection operator $\otimes$, the hypothesis that $\mu_{\tilde{S}_1 \cap \tilde{S}_2}(\omega) = \otimes\big(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\big)$ is empirically adequate, supposing $\mu_{\tilde{S}_1}(\omega)$ and $\mu_{\tilde{S}_2}(\omega)$ to be determined via the measurement procedure associated with the operational definition of graded membership.[2]

To answer this question, we started by bringing together the results from the aforementioned empirical work concerned with specific atomic concepts. We then derived from those results what the various intersection operators predict for a concept combining the atomic concepts. Finally, we conducted a new study to elicit people's judgments about the degrees to which various items fall under the combined concept, in order to compare those with the predicted degrees.[3]

Section 2 describes the theoretical background of the results to be presented in this paper, in particular, the conceptual spaces framework and the account of graded membership it gives rise to. Section 3 summarizes the earlier experimental work whose outcomes were used to make the predictions that the study to be reported in Section 4 aimed to test. Finally, Section 5 discusses the broader implications of the results from the new study for the status of fuzzy set theory.

2 Theoretical background

We first summarize the conceptual spaces framework, which serves as the backdrop for all that follows. We then describe a recent proposal for modeling vagueness in that framework. And finally we explain how this model for vagueness can be used to define graded membership.

2.1 The conceptual spaces framework

Over the past fifty years, cognitive psychologists have developed a mathematical framework for modeling concepts *geometrically*, as regions in so-called similarity spaces. Similarity spaces are one- or multi-dimensional mathematical structures with a metric defined on them. The dimensions of a similarity space represent fundamental qualities that items can have to a greater or lesser degree, so that items can be mapped onto points in the space, according to the degrees to which they have the given qualities. Distance in the space is meant to represent dissimilarity, in that the closer two items are in the space, the more similar they are to each other in the respect modeled by the space (e.g., in respect to color, if the space happens to be color space—see below).[4]

[2]More exactly, the question is whether this holds, *given an application context*. As is generally acknowledged, fuzzy set theory is sensitive to context (or "local," in Zadeh's 1975 terminology). More on this below.

[3]By proceeding in this way, we have effectively answered a call by Bělohlávek, Dauben, and Klir (2017, p. 443), who stress the need for more experimental work on the notion of graded membership, both in itself and in relation to various set-theoretic operations. See in this vein also Bělohlávek and Klir (2011a, p. 261).

[4]For a detailed presentation of the conceptual spaces framework, see Gärdenfors (2000). See Gärdenfors (2014), Douven (2019a), and Douven and Gärdenfors (2020) for some amendments and refinements.

Because in the theoretical work to be summarized, the conceptual spaces framework was invoked to provide a solid footing for the notion of graded membership, it is worth briefly going into the provenance of similarity spaces. The input data for the construction of such spaces are either similarity judgments or so-called confusion probabilities (basically, data concerning the likelihood that two distinct stimuli are mistakenly judged to be identical when consecutively flashed to participants). Such data are then standardly transformed into distances, whereupon a multi-dimensional scaling (MDS) procedure or, less frequently, a variant thereof (e.g., non-negative matrix factorization) is conducted to obtain a metrical space.

In an MDS procedure, the aim is to arrive at a space that, first, is *low-dimensional*, ideally with at most three dimensions; second, has *good fit*, standardly measured in terms of "stress," with lower stress values indicating more faithful representations of the similarities or confusion probabilities that served as input data; and third, has *interpretable dimensions*, meaning that each dimension can be linked to some fundamental quality the stimuli can be said to have to some varying degree (Borg & Groenen 2000).

It is to be emphasized that there is no guarantee that the outcome of an MDS procedure will satisfy all, or even just some, of these three criteria. But if it does, we still only have a similarity space, and not a conceptual space. There are different ideas about how similarity spaces can come to represent concepts, but the dominant one relies on the notion of prototype together with the mathematical technique of Voronoi tessellations. According to *prototype theory*, some instances of a concept are more representative for it than others, and its most representative instance is the concept's prototype (Rosch 1973, 2011). Prototypes are known to play various central roles in cognition. For instance, they are typically recognized faster than non-prototypical instances and they are also more easily memorized. A *Voronoi tessellation* of a given space is a partitioning of that space into disjoint cells such that each cell is associated with exactly one so-called generator point and contains all and only those points in the space that lie at least as close to that cell's generator point as they lie to any other cell's generator point; see the left panel of Figure 1 for an illustration and see Okabe et al. (2000) for formal details. The points lying equally close to two or more generator points constitute the boundaries between cells. The idea, then, is that we obtain a conceptual space from a similarity space by locating, within the latter, the prototypes of various concepts, and using these to generate a Voronoi tessellation of the space, whose cells then form the concepts represented by the space.

A major advantage of this approach is that it makes the study of concepts formally rigorous as well as empirically testable. For example, we can now conceive the concept of redness as a region in CIELUV space (see below), which means we can carry out all sorts of mathematical operations on it—like measuring its volume—and at the same time use it for conducting all sorts of experimental work. Indeed, various empirical consequences of this approach—such as, most famously, that concepts form convex regions in a similarity space—have already been borne out by observations; see Gärdenfors (2000, Ch. 3).

2.2 Modeling vagueness

One of the key research questions in the literature on vagueness concerns the metaphysical status of borderline cases, where a borderline case of a given concept C is generally characterized as one that is neither clearly C nor clearly not C (e.g., Wright 1992; Williamson 1994; Égré & Barberousse 2014). When seen through the lens of the conceptual spaces framework, however, that question appears to have a straightforward answer. After all, we just said that points which lie maximally close to two or more prototypical points (i.e., points in the relevant space that represent prototypes) constitute

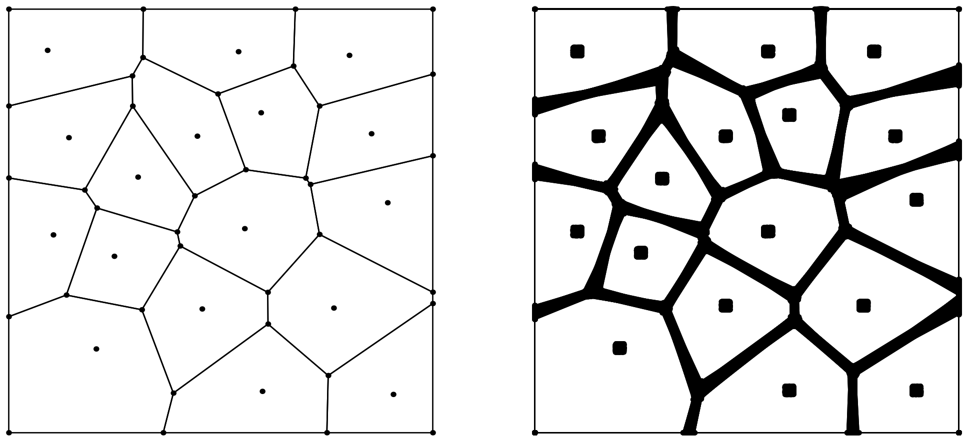

Figure 1: The left panel shows one of the (ordinary) Voronoi tessellations that make up the collated Voronoi tessellation shown in the right panel.

the boundaries between cells, and hence represent the boundaries between concepts. So, why not simply say that a borderline case of a concept C is a case whose representation in the space in which we represent C lies equidistantly between the C-prototypical point and at least one other prototypical point in that space?

While this answer is simple and seemingly natural, Douven et al. (2013) argue that it is not completely satisfactory. As these authors point out, one can have, say, blue/green borderline cases such that any shade just slightly different from them counts again as a blue/green borderline case. But this observation would be difficult to account for if the boundaries between concepts, and thus also between blue and green (which are adjacent in color space), were really only "one point thick," as they are in the standard conceptual spaces framework. For if that were the case, then "almost all" (in the measure-theoretic sense of this expression) ever-so-slight departures from the color of a blue/green borderline case would result in a clear case of either blue or green, given that any point on the boundary between concepts is immediately next to points *not* belonging to that boundary.

Douven et al. (2013) propose a solution to this "thinness" problem that starts from the observation that, for many concepts, it is actually wrong to assume that the concept has precisely one prototype. Try to imagine, for instance, a shade that strikes you as typically blue, and now imagine a shade that just slightly differs from that. (If you have difficulty imagining this, you can use a drawing program of the kind standard on most computer and first select a color that you deem typically blue, and then just slightly alter one of the parameters that the program uses for defining colors.) Chances are that the resulting color will again strike you as a typical instance of blue. There is also empirical support for the non-uniqueness of prototypical instances. Notably, Berlin and Kay (1969) found that the participants in their famous color-naming experiment, when asked which of the Munsell chips used in the experiment they regarded as typical for a given color, often pointed at several of them; for further evidence, see Section 3. Taking on board the idea that, frequently, prototypes are non-unique, Douven et al. (2013) proposed that conceptual spaces are to be equipped with prototypical *regions* instead of prototypical *points*.[5]

[5]In Douven et al.'s (2013) proposal, these regions have certain elementary properties (e.g., they are required to be connected in the topological sense of the word); such formal details need not detain us here. For an interesting *probabilistic* interpretation of prototypical regions, see Verheyen and Égré (2018).

But then, what remains of Voronoi tessellations, the other key component of the conceptual spaces framework, given that they are defined only for sets of single generator points? To address this question, Douven et al. (2013) proposed a modification of the technique of Voronoi tessellations. In this modification, one takes the set of all possible selections of one single point from each prototypical region in a space, lets each of those selections generate a Voronoi tessellation of the space, and then "superimposes" all the tessellations thus obtained onto each other, thereby constructing a (what they call) *collated* Voronoi tessellation of the space. The right panel of Figure 1 shows an example of a collated Voronoi tessellation of a two-dimensional space. It is immediately obvious how this construction solves the thinness problem: concept boundaries are now *thick*, so that "almost all" borderline cases are surrounded by other borderline cases.

It would seem, however, that even in the new picture not all is well. After all, that picture still suggests that there is sharp line between the cases *clearly* falling under a concept and that same concept's borderline cases. And as various authors have argued, the distinction between clear cases and borderline cases appears to be vague itself. (This is often referred to as second-order vagueness; see, e.g., Wright 1992 or Zardini 2013.) This remaining issue was addressed by the account of graded membership developed in the conceptual spaces framework with the collated Voronoi tessellations add-on. We first describe this account and then come back to the question of second-order vagueness.

2.3 From vagueness to graded membership

Decock and Douven (2014) propose to use the extended conceptual spaces framework described above for defining graded membership.[6] As said previously, their proposal consists of providing an operational definition of, and thus a method for *measuring*, the degrees to which items fall under a given concept. The measure they put forward is easiest to describe for the kind of case in which each prototypical region in a space has only finitely many points. Consider, for example, the two-dimensional space depicted in Figure 2. That space contains four prototypical regions, each of which consists of two points: $\{a, b\}$, $\{c, d\}$, $\{e, f\}$, and $\{g, h\}$. There are $2^4 = 16$ ways to select one point from each prototypical region, and so the collated Voronoi tessellation on this space will consist of 16 superimposed ordinary Voronoi tessellations. In Decock and Douven's proposal, a point in this space belongs to a concept to a degree that equals the number of those ordinary tessellations that locate the point in the cell associated with one of the two prototypical points associated with the concept, divided by the total number of ordinary Voronoi tessellations. For instance, with some effort one can see in the figure that 8 of the 16 ordinary Voronoi tessellations that make up the collated Voronoi tessellation on this space put the point i in the cell associated with a or b. Hence, i falls under the concept with prototypical region $\{a, b\}$ to a degree of $1/2$. One similarly verifies that j falls under the same concept to a degree of $1/4$.

This idea generalizes swiftly to all conceptual spaces whose prototypical regions are constituted by finitely many points. More importantly, it generalizes further to conceptual spaces some or all of whose prototypical regions consist of infinitely many points. This generalization is more intricate, and readers are referred to Decock and Douven (2014) or Douven (2019b) for mathematical details. In brief, the generalization starts from the observation that the ordinary Voronoi tessellations making up the collated Voronoi tessellation of a space are themselves representable as points in a different,

[6] Decock and Douven (2014) built on important work by Kamp and Partee (1995), which in turn was a continuation of work begun in Kamp (1975). In particular, Douven and Decock showed how the questions concerning graded membership that Kamp and Partee had to leave open all received and obvious and natural answer once those authors' proposal was embedded in a conceptual spaces framework equipped with collated Voronoi tessellations.

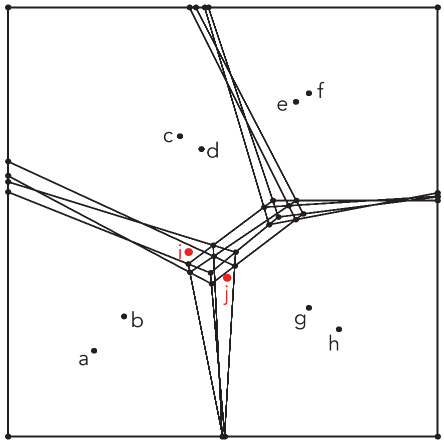

Figure 2: Collated Voronoi tessellation whose prototypical regions contain only finitely many points.

more abstract, space. Suppose, for instance, we are considering an m-dimensional space S with n prototypical regions in it. Then the ordinary Voronoi tessellations generated by the said kind of selections of points from the prototypical regions can be represented by points in an $m \times n$-dimensional space S^*: for each tessellation, one chooses n points, each represented by m coordinates. Then, given a concept C representable in S, the degree to which a given item x falls under C can be defined as the Lebesgue measure of the set of points in S^* which represent Voronoi tessellations that group the point representing x in S together with the clear C cases in S divided by the Lebesgue measure of S^* as a whole. It can easily be shown that, given this definition, a concept C's clear cases all have a degree of C-membership of 1, its clear non-cases all have a degree of membership of 0, and its borderline cases all have an intermediate degree of membership. It can further be shown that, for borderline cases, degree of membership tapers off in the form of an S-shaped curve, from 1 to 0, as we move away from the concept's prototypical region.

Importantly, on this account degree of membership is not simply a function of distance from the prototype or prototypes, an idea that Osherson and Smith (1981), for instance, had mistakenly assumed fuzzy logic to be committed to and which they then went on to criticize on the grounds of the in itself correct observation that atypical instances of a concept can still clearly fall under that concept (e.g., an ostrich is clearly a bird, even if an atypical one). Another look at the right panel of Figure 1 or at Figure 2 will suffice to appreciate that the present account actually *predicts* the existence of cases fully falling under a concept, yet being at different distances from that concept's most typical instances. For instance, an ostrich would probably be represented by a point relatively far from the prototypical bird region in the appropriate conceptual space, but it would still lie within the region of clear bird cases.[7]

Note also how the fact that, in the boundary region, degree of membership tapers off *gradually* helps to solve the problem of second-order vagueness mentioned previously. The concern was that, in reality, we experience a smooth transition from clear cases to borderline cases to clear non-cases, whereas the collated Voronoi tessellation shown in the right panel of Figure 1 suggests that, in our model, there cannot but be an abrupt transition if we go from clear cases to borderline cases and equally if we go from borderline cases to clear non-cases. But the above account of graded membership allows

[7] For a more systematic critique of Osherson and Smith's 1981 paper, see Bělohlávek and Klir (2011b).

us to argue that we experience these transitions as smooth because they *are* smooth. While, in our model, there are still sharp cut-offs between the different types of cases (i.e., clear cases, borderline cases, and clear non-cases), the differences in degree of membership at these cut-off points are infinitesimally small, which explains why we fail to feel any kind of jolt when moving across the lines. For a more extensive discussion of this point, see Douven and Decock (2017, Sect. 3).

Finally, it merits emphasis how the present account is able to deal with a criticism of fuzzy set theory that is particularly common among philosophers, to wit, the claim that this theory entails an artificial precision by assuming that, for instance, for a given borderline blue/green case, there is a *specific* degree to which it is blue; it just strains credulity—critics like Haack (1979) and Keefe (2000) hold—that the case is blue to a degree of exactly r, with $r \in [0, 1]$. However, as Gärdenfors (2000, p. 89) has argued on independent grounds, the conceptual spaces approach is idealized in that psychological measures in general, and the metrics defined on conceptual spaces in particular, tend to be inexact. For the purposes of modeling, we may assume that a conceptual space has a precise structure, but in practice people will be uncertain about this structure. As a result, "the borderline [of a Voronoi tessellation on a conceptual space] will not be exactly determined" (*ibid.*). This indeterminacy about the locations of the borderlines in single Voronoi tessellations carries over to the boundary regions of collated Voronoi tessellations that, after all, are made up of single tessellations (Douven et al. 2013, p. 153), and thence to the degrees of membership of items representable in or very near to such a boundary region. In light of this, we should not expect that it will generally "feel" to us as though a given borderline case of a concept has an exact degree of falling under that concept. Naturally, that the conceptual spaces framework, and thereby also the account of graded membership that builds on it, involves some idealization can hardly be deemed objectionable, lest we should regard most scientific theories as being flawed for relying on idealizing assumptions.

3 Relevant empirical results

As mentioned in the introduction, fuzzy set theory was motivated by the observation that some things appear to not quite belong to a given set but also appear to not be quite outside that set. By providing an operational definition of graded membership, we have begun to address the concern that fuzzy set theory has been built on sand. This is only a *beginning* because it still remains to be seen whether, if it appears to us that x is a C to a degree of somewhere around y, then that is also what follows from our model, and vice versa. If fuzzy set theory, when combined with our operational definition, turns out to be empirically inadequate, then the theoretical work reported in the previous section will have been all for naught.

Thus arises the question of how we could test our operational definition. Here is an apparently straightforward recipe: Find a conceptual space whose geometrical structure is known or can be determined, locate the prototypical regions in that space, and then, assuming our definition, compute the degrees to which various items belong to the concepts represented by the space. As a next step, let people judge the degrees to which the given items fall under the concepts. Finally, compare those judgments with the predicted degrees.

We face a hurdle right at the first step. Although there is much published research about a variety of conceptual spaces (for examples and references, see, e.g., Gärdenfors 2000, 2014), one is hard pressed to find data that would allow one to *work* with those spaces. Color space is, to some extent, an exception. There are actually two spaces generally considered to adequately represent human judgments of color similarity, to wit, CIELAB space and CIELUV space. Both are three-dimensional Euclidean spaces,

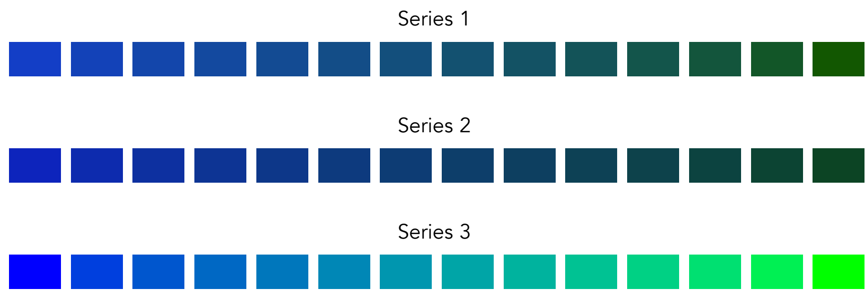

Figure 3: The three color series that served as part of the materials for the experiments reported in Douven et al. (2017).

and they are very similar in shape. Nevertheless, it has been experimentally established that CIELAB space is the better choice to work with when the colors whose similarities are to be judged are presented on cloth or on paper, while CIELUV space is preferable when the colors are perceived on a screen (Malacara 2002; Fairchild 2013). Various software packages, such as MATLAB and *Mathematica*, make it easy to visualize these spaces and also to perform various operations in them (to measure distances, volumes, etc.). Unfortunately, this still only gives us access to the geometry of those spaces; information about where, in these spaces, we find the prototypical regions of colors—even just of the so-called basic colors—is still sparse.[8] Thus, at a minimum, we will first have to determine the locations in color space of the prototypical regions before we can calculate to what extent a given shade is red, or blue, or yellow, and so on.

Douven et al. (2017) made a beginning with locating the prototypical regions in color space, specifically in CIELUV space (given that their experiments were conducted on-line). Their studies were limited to the prototypical blue and prototypical green regions. First, they showed participants a great number of cross-sections of RGB space, asking them to click on those points in the cross-sections that appeared most typically blue to them, and in another part of the experiment, to click on those points that appeared most typically green to them, where it was randomly determined which part a participant would be shown first, and where also the cross-sections appeared in an order that was randomized per participant. Douven and colleagues used the results from this study to identify the approximate regions in CIELUV space where shades were deemed prototypically blue and green, respectively, and sampled from those regions a great variety of shades that were then presented to participants in a second study, where these participants were now asked to indicate for each of the sampled shades whether they were typically blue or green, respectively. Finally, the convex hulls of the sets of those shades that received sufficiently many positive responses were assumed to give a reliable indication of where in CIELUV space the prototypical blue and green regions are to be found.

Douven and coauthors' primary interest was not in the structure of color space per se, but in testing the operational definition of graded membership discussed above. Knowledge of that structure of color space was needed to calculate the degree of blueness/greenness of various color

[8] It is known for a small subset of the Munsell chips which of those chips are thought to show prototypical colors. The information this gives is limited, however, inasmuch as the chips in the subset all lie on the surface of color space (whether CIELAB space or CIELUV space). Systematic research into the "deeper" layers of color space has only just begun; see Jraissati and Douven (2018) and Douven (2019c). Nonetheless, already work reported in Douven et al. (2017) shows that not all prototypical colors lie on the surface of color space.

shades, specifically all the shades occurring in Series 1–3 shown in Figure 3. It was a further part of their study, then, to elicit from participants judgments of the degrees of membership in the set of blue/green things for each of those shades. Figure 4 shows the locations of those shades as well as the locations of the prototypical blue and green regions in CIELUV space.[9]

Because the shades in Series 1–3 are unmistakably in the blue–green region of color space, Douven and coauthors assumed that if the locations in CIELUV space of prototypical regions other than for blue and green were known, this would not lead to degrees of blueness/greenness of the shades in the series appreciably different from those obtained by applying the operational definition of graded membership just to CIELUV space equipped with the prototypical blue and green regions. Hence, they proceeded on the basis of knowledge of the prototypical blue and green regions, calculated on this basis the degrees of blueness/greenness of the shades in Series 1–3, and compared the outcomes with the judgments about the degrees of blueness and greenness of those shades that they had elicited from participants in their study.

It was said that, in Douven et al. (2017), the prototypical blue and green regions were determined by taking the convex hulls of the sets of shades that had received sufficiently many "typically blue" and "typically green" responses, respectively. Realizing that there was some arbitrariness in setting a threshold for what is to count as "sufficiently many," Douven and colleagues in effect considered a number of possible thresholds, to wit, every quartile of highest-rated shades, and also all shades that made the cut in the first experiment. They then predicted degrees of blueness and greenness for the shades used in Series 1–3 on the basis of each of those thresholds separately, and they compared the observed degrees with all of the thus obtained predictions.

We see a graphical presentation of both the predicted degrees (for the various thresholds) and the observed (i.e., judged) degrees in Figure 5. From the first row, it is obvious that the choice of threshold is largely immaterial: they all lead to basically the same predictions. The second and third rows show that predicted and observed degrees of blueness match closely for all three series, even with a near-to-perfect match for the first series. Probably, the most important observation is that the

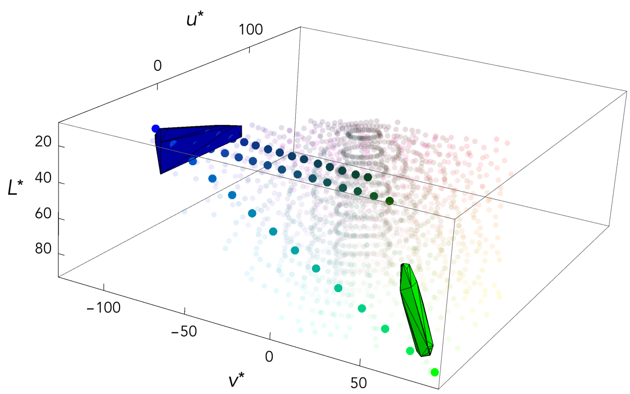

Figure 4: CIELUV space with Series 1–3 and prototypical regions for blue and green. Series 3 is the one with the greatest span and Series 2, which appears as the "uppermost" series, the one with the shortest span.

[9]The figure shows the global structure of CIELUV space by thinly representing the full set of Munsell chips in the space; see Douven et al. (2017) and Jraissati and Douven (2018) for details.

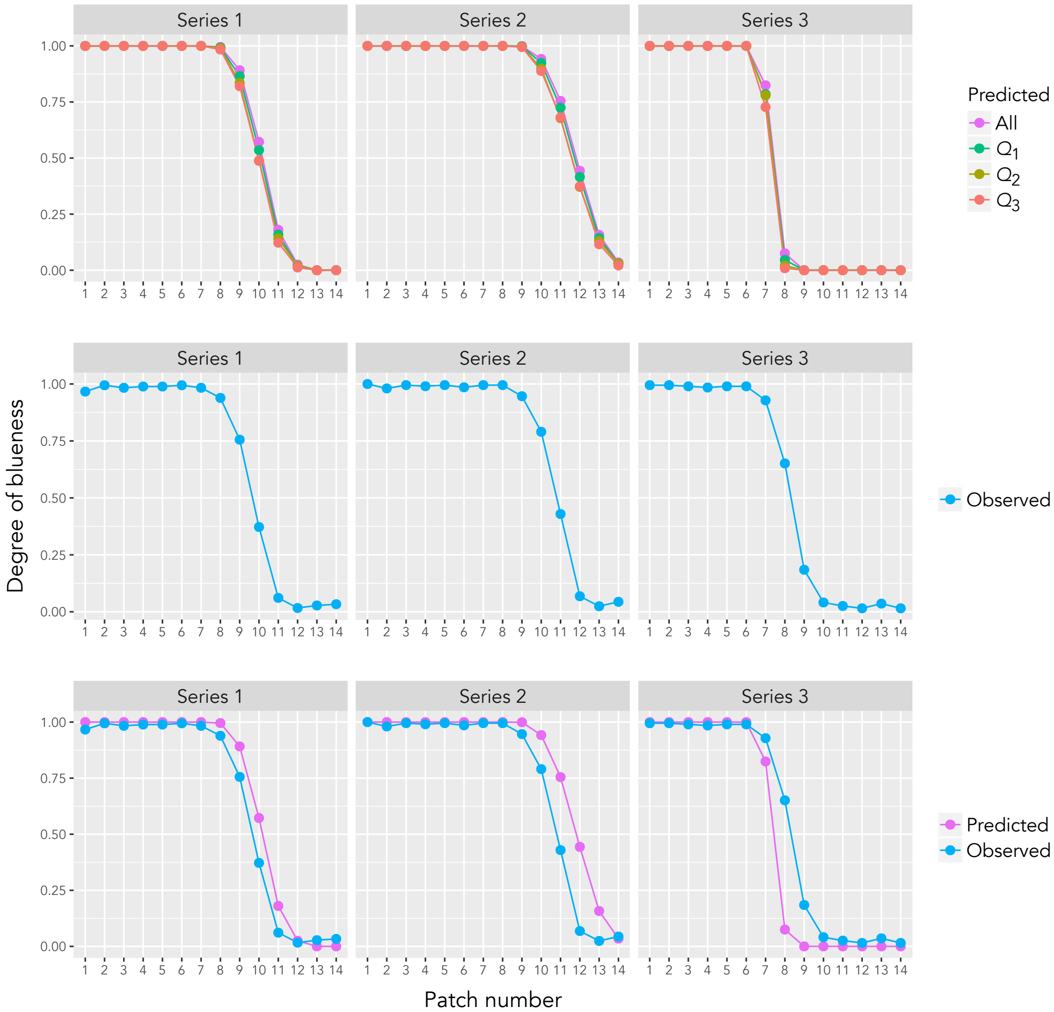

Figure 5: Top row showing predicted degrees of blueness for the patches in Series 1–3, assuming different thresholds for prototypicality, with Q_i giving the degrees of blueness assuming as prototypical regions the convex hulls of the points with numbers of clicks greater than or equal to the i-th quartile (and "All" being the totality of clicks; see the text for explanation). Middle row showing observed degrees of blueness. Bottom row showing predicted versus observed degrees of blueness (with predictions based on all clicks).

predicted degrees for any of the three series are closer to the observed degrees for that same series than they are to the observed degrees for either of the other series. And finally, we see that, for each series, the graph of the membership function is clearly S-shaped, which was a general prediction of Decock and Douven's (2014) definition of graded membership.[10]

In the statistical analysis of their data, Douven and coauthors started by fitting logistic curves to the observed as well as to the predicted degrees for each series and then determined, for each curve, its

[10] This was also expected in light of the findings in work on vagueness and graded membership by other authors (e.g., McCloskey & Glucksberg 1978; Smith & Minda 1998; Hampton 2007). See Douven (2016b, Sect. 7) and Douven et al. (2017, Sect. 7) for a comparison of that work with the work discussed here.

so-called Point of Subjective Experience (PSE; basically, the point at which the curve crosses the line $y = 0.5$). Finally, they calculated the slope of each curve at its PSE. The outcomes of this analysis were as expected, given the graphs in Figure 5: matches between predictions and observations were generally excellent and were significantly better within series than across series. A comparison of sums of squared deviations of predictions from observations led to the same conclusion.

Given that these results concerned only color space, and then even just the blue–green region in that space, they only yielded *some* confirmation of Decock and Douven's proposal for defining graded membership operationally. At the same time, however, the work provided a kind of blueprint for further experiments, which could perhaps focus on larger parts of color space or on different spaces altogether. One such follow-up study is presented in Douven (2016b), which looks at graded membership in a space for the representation of container-like objects. This work was more involved than the work reported in Douven et al. (2017), given that, where the latter used a space whose geometrical structure was known, the container-shape space used in Douven (2016b) had to be constructed from scratch. For that reason, the work had to start with conducting a study that would provide the data needed for building a similarity space. Once that been accomplished, a study was conducted to locate the relevant prototypical regions in the space. At that point, degrees of

Figure 6: The 49 figures that were used as stimuli in all experiments reported in Douven (2016b). The numbers in the bottom-right square of each grid were not shown to participants; they serve here as labels to facilitate interpretation of Figure 7.

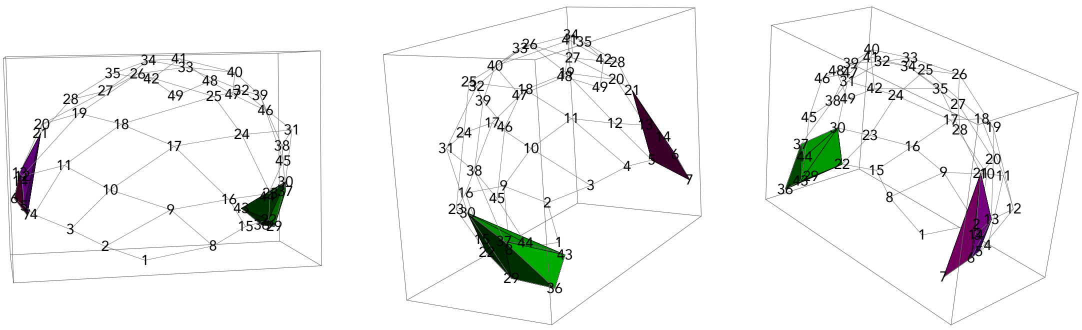

Figure 7: Three viewpoints of the three-dimensional city-block space for container-like shapes, with convex hulls of majority choices of typical bowl shapes (in green) and typical vase shapes (in purple); the numbers in the space refer to the shapes in Figure 6.

membership for the relevant container-shaped items could be calculated. And finally, a study was conducted to elicit human judgments of the degrees of membership for those items, which in the analysis could then be compared with the predicted degrees.

The studies all used as stimuli the 49 shapes shown in Figure 6. The first study involved more than 1,000 participants, each participant being shown 25 randomly chosen pairs of stimuli. Participants were asked to rate the similarity of the members of each pair. The similarity ratings were aggregated across participants and the aggregates served as the input data for an MDS procedure. From this procedure the three-dimensional city-block space shown in Figure 7 emerged as doing best across all model-fit criteria; in particular, the associated stress value was exceedingly low, and the dimensions of the space all had natural interpretations.[11]

Again, this yielded only a similarity space, and not a conceptual space. To obtain that, the locations of relevant prototypical regions still had to be determined. Prima facie it seemed that concepts that could plausibly be interpreted in this space included, most notably, those of bowl, cup, mug, pot, and vase. However, a prestest showed that there was little support for the thought that any of the shapes in Figure 6 is typical for cups, mugs, and pots. On the other hand, a majority of participants did judge various shapes to be typical for bowls, and various other shapes to be typical for vases. Figure 7 shows where these shapes are to be found in the three-dimensional city-block space, coloring the convex hull of the typical bowl shapes green and that of the typical vase shapes purple.

At this point, degrees of bowlhood and vasehood of the 49 stimuli could be computed, again assuming Decock and Douven's operational definition of graded membership. These degrees were then compared with the degrees of vasehood and bowlhood for those stimuli as judged by over 500 participants of a final study in Douven (2016b). In this study, degrees of bowlhood and vasehood were elicited in a variety of ways (this was done for control purposes; see Douven 2016b for details). The analysis then showed that, however the degrees were elicited, there was an excellent match, assuming any pertinent goodness-of-fit measure, between predicted and observed degrees. Figure 8 presents predicted and observed degrees graphically, which also gives a good impression of the close fit.

The figure is to be interpreted as follows: The leftmost points of the green and purple surfaces—so where the green surfaces are close to 0 and the purple surfaces close to 1—correspond to the left

[11] The space shown in Figure 7 is a *city-block* space because distances in the space are measured by the city-block metric (also known as "Manhattan metric" or "taxicab metric"). This is to say that, for any pair of points (v_1, v_2, v_3) and (w_1, w_2, w_3) in the space, their distance is defined to be $|v_1 - w_1| + |v_2 - w_2| + |v_3 - w_3|$.

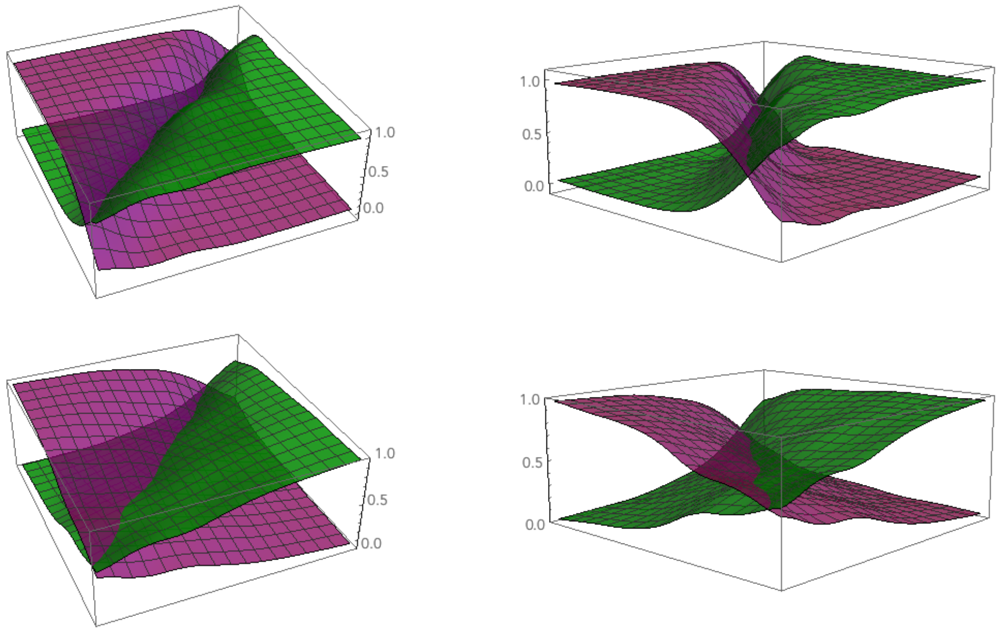

Figure 8: Different views on the membership functions for bowl (green) and vase (purple) for predicted (top row) and observed (bottom row) degrees of bowlhood and vasehood. See the text for explanation.

upper point in Figure 6, specifically stimulus 7 in that figure; the rightmost points—where the green surfaces are close to 1 and the purple surfaces close to 0—correspond to the right lower point in Figure 6, specifically stimulus 43. Imagine the surfaces in Figure 8 to be projected onto Figure 6 in that way; that gives an impression of the degrees of membership accrued by the various stimuli. It is readily observed that what came out as the prototypical bowl and vase shapes in one of the studies (see again Figure 7) got the highest degrees of membership also according to the observations. It is equally clear that the graphs of the membership functions are S-shaped again (specifically S-shaped manifolds). Douven (2016b) notes that the predicted degrees slightly over-accentuate the S-shapes of the membership functions, in comparison with the observed degrees. As he points out, however, that may well be a central tendency effect, which tasks involving Likert scales (which were used for eliciting participants' judgments concerning the degrees of bowlhood and vasehood) are known to give rise to (see Douven 2018 and references given there).

4 Study

The empirical results from previous studies, summarized in Section 3, vindicate the operational definition of graded membership as presented by Decock and Douven, a definition aimed at putting to rest concerns about the very foundations of fuzzy set theory. It is to be noted, however, that these results do not *ipso facto* vindicate fuzzy set theory itself. It is one thing to show that the notion a theory builds on is meaningful and bona fide, and it is another thing to show that the theory built on that notion is any good, or at any rate that it is empirically adequate, as Zadeh intended it to be. In particular, the earlier results do nothing to demonstrate the empirical adequacy of the fuzzy extensions of the classical complement, inclusion, union, and intersection operations.

In this paper, we concentrate on the various fuzzy intersection operators that have been forwarded. As Zimmermann (2001, p. 43) notes, it is somewhat confusing that there are so many of them. Which one, if any, are we to use? Zimmermann is in fact skeptical that we will be able to identify a unique intersection operator that will be usable in all application contexts. In his opinion, the best we can hope for is to have "a very small number of operators to model many situations" (*ibid.*).

Zimmermann (*ibid.*) further observes that there is no a priori way to settle the question of which operator or operators to go with: "If fuzzy set theory is used as a modeling language for real situations or systems, . . . the operators must be appropriate models of real-system behavior; and this can normally be proven only by empirical testing." In line with this observation, we aim to provide the question of which intersection operator or operators to prefer with an empirical answer. In doing so, we want to keep on board the operational definition of graded membership that was the focus of previous empirical work. If we were somehow able to buttress by data a specific intersection operator or specific class of such operators, but only at the expense of sacrificing the operational definition of graded membership, we would be back at the point where fuzzy set theory could be challenged for relying, at its base, on a mysterious notion of things sometimes belonging to sets only *partially*. Given the way the operational definition depends on the conceptual spaces framework, investigating the empirical adequacy of intersection operators is only possible if we can avail ourselves of some conceptual spaces. In the previous section, we presented two conceptual spaces that are accessible to anyone who has the appropriate software installed on their computer. Those are precisely the spaces that were used for our current research purposes.

We can state our specific research questions, as follows:

Q1. For items that are

   (i) representable in the shape space from Douven (2016b); and

   (ii) appear in a shade representable in the blue–green region of CIELUV space,

   does Decock and Douven's account of graded membership yield accurate predictions for the degrees to which they are judged to be blue vases by people?

Q2. If so, can we obtain accurate predictions from that account using the degrees of blueness and degrees of vasehood as they can be derived from the structures of the aforementioned spaces?

Q3. If so, can we obtain accurate predictions from that account by combining degrees of blueness and degrees of vasehood via one of the known fuzzy intersection operators?

Note that, beforehand, there is no reason to hold that Q1 can be answered in the positive. The operational definition has so far only been tested for atomic concepts; it might break down as soon as we start to combine concepts. As for Q2, even if the answer to Q1 is positive, it is still the case that, for all we know, some new conceptual space has to be constructed, on the basis of similarity judgments concerning the shapes shown in Figure 6 but now appearing not in neutral gray but in various shades of blue and green. Only that space might—*might*, because even this is not guaranteed—make empirically accurate predictions about the degrees to which people would deem the items blue vases.[12] And supposing Q2 *can* be answered in the positive, it is a further open question whether any

[12] Is it not already clear that at least for some conjunctive concepts, the question analogous to Q2 will have a negative answer (even supposing the question analogous to Q1 has a positive answer)? A referee pointed out that the infamous concept of being a pet fish may present special difficulties when it comes to deriving degrees of membership for that concept

of the known fuzzy intersection operators will help to derive accurate predictions; perhaps we must look beyond those. We now briefly describe the intersection operators that were considered in our study.

4.1 Fuzzy intersection operators

The fuzzy intersection operators to be looked at can be divided into averaging operators and triangular norms (or "t-norms"), where the latter can again be divided into parameterized and non-parameterized norms. For a function to be a t-norm it must satisfy the properties of monotonicity, commutativity, and associativity (see, e.g., Dubois & Prade 1980; Zimmermann 2001, Ch. 3); the averaging operators do not satisfy all these properties. The parameterized t-norms, and also the averaging operators, take as input not only degrees of membership but also an additional parameter; the non-parameterized t-norms have only degrees of membership as input.

Below are the best-known non-parameterized t-norms:

$$
\begin{aligned}
t_{\mathrm{M}}\left(\mu_{\tilde{S}_1}(\omega),\mu_{\tilde{S}_2}(\omega)\right) &:= \min\left[\mu_{\tilde{S}_1}(\omega),\mu_{\tilde{S}_2}(\omega)\right]\\
t_{\mathrm{P}}\left(\mu_{\tilde{S}_1}(\omega),\mu_{\tilde{S}_2}(\omega)\right) &:= \mu_{\tilde{S}_1}(\omega)\cdot\mu_{\tilde{S}_2}(\omega)\\
t_{\text{Ł}}\left(\mu_{\tilde{S}_1}(\omega),\mu_{\tilde{S}_2}(\omega)\right) &:= \max\left[0,\mu_{\tilde{S}_1}(\omega)+\mu_{\tilde{S}_2}(\omega)-1\right]\\
t_{\mathrm{D}}\left(\mu_{\tilde{S}_1}(\omega),\mu_{\tilde{S}_2}(\omega)\right) &:= \begin{cases}\min\left[\mu_{\tilde{S}_1}(\omega),\mu_{\tilde{S}_2}(\omega)\right] & \text{if } \max\left[\mu_{\tilde{S}_1}(\omega),\mu_{\tilde{S}_2}(\omega)\right]=1\\ 0 & \text{otherwise}\end{cases}\\
t_{\mathrm{H}}\left(\mu_{\tilde{S}_1}(\omega),\mu_{\tilde{S}_2}(\omega)\right) &:= \begin{cases}0 & \text{if } \mu_{\tilde{S}_1}(\omega)=\mu_{\tilde{S}_2}(\omega)=0\\ \frac{\mu_{\tilde{S}_1}(\omega)\cdot\mu_{\tilde{S}_2}(\omega)}{\mu_{\tilde{S}_1}(\omega)+\mu_{\tilde{S}_2}(\omega)-\mu_{\tilde{S}_1}(\omega)\cdot\mu_{\tilde{S}_2}(\omega)} & \text{otherwise}\end{cases}\\
t_{\mathrm{E}}\left(\mu_{\tilde{S}_1}(\omega),\mu_{\tilde{S}_2}(\omega)\right) &:= \frac{\mu_{\tilde{S}_1}(\omega)\cdot\mu_{\tilde{S}_2}(\omega)}{2-\left(\mu_{\tilde{S}_1}(\omega)+\mu_{\tilde{S}_2}(\omega)-\mu_{\tilde{S}_1}(\omega)\cdot\mu_{\tilde{S}_2}(\omega)\right)}
\end{aligned}
$$

The first of these, the minimum t-norm t_{M}, is in effect Zadeh's (1965) definition of fuzzy intersection. The others are, respectively, the product t-norm (also to be found in Zadeh 1965), the Łukasiewicz t-norm (see Borgwardt, Cerami, & Peñaloza 2017), the drastic product (Dubois & Prade 1980), the Hamacher product (Hamacher 1978), and the Einstein product (Mizumoto & Tanaka 1981). It is easy to show that these t-norms obey pointwise the following ordering: $t_{\mathrm{D}} \leqslant t_{\text{Ł}} \leqslant t_{\mathrm{E}} \leqslant t_{\mathrm{P}} \leqslant t_{\mathrm{H}} \leqslant t_{\mathrm{M}}$ (Zimmermann 2001, p. 32).

We mentioned that, according to Zimmermann, it would be overly ambitious to aim for a one-size-fits-all intersection operator, if that operator is to be applied to fuzzy sets; context—the type of objects of which sets are intersected—may matter. Other authors were of the same opinion. Thus were proposed a number of parameterized t-norms, the idea being that these could be tailored for

from degrees of membership for the constituent concepts. But while the concept of being a pet fish has been considered problematic for fuzzy set theory in general (see, e.g., Osherson & Smith 1981), by relying on the conceptual spaces framework we may be well-positioned to tackle it, given how the framework is able to handle the effects of context and contrast classes on concepts. For instance, in the context of skin colors, color concepts are to be thought of as being represented in a particular *subspace* of color space, meaning that the concept of yellowness, for instance, is different, dependent on whether we are discussing the color of flowers (say) or skin colors (Gärdenfors 2000, p. 121 f). Similarly, the concept of being a pet may be different when discussed in connection to fish than when discussed in connection to cats or dogs. See Gärdenfors (2000, Ch. 4) for an extensive discussion of these matters.

each context separately by setting a parameter to a specific value. The parameterized t-norms to be considered in the following are these:

$$t_{\mathrm{DP}}^{\gamma}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) := \frac{\mu_{\tilde{S}_1}(\omega) \cdot \mu_{\tilde{S}_2}(\omega)}{\max\left[\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega), \gamma\right]}, \quad 0 \leqslant \gamma \leqslant 1$$

$$t_{\mathrm{F}}^{\gamma}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) := \begin{cases} t_{\mathrm{M}}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) & \text{if } \gamma = 0 \\ t_{\mathrm{P}}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) & \text{if } \gamma = 1 \\ t_{\mathrm{Ł}}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) & \text{if } \gamma = \infty \\ \log_{\gamma}\left(1 + \frac{\left(\gamma^{\mu_{\tilde{S}_1}(\omega)} - 1\right)\left(\gamma^{\mu_{\tilde{S}_2}(\omega)} - 1\right)}{\gamma - 1}\right) & \text{if } \gamma > 0, \gamma \neq 1, \gamma \neq \infty \end{cases}$$

$$t_{\mathrm{H}}^{\gamma}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) := \frac{\mu_{\tilde{S}_1}(\omega) \cdot \mu_{\tilde{S}_2}(\omega)}{\gamma + (1 - \gamma)\left(\mu_{\tilde{S}_1}(\omega) + \mu_{\tilde{S}_2}(\omega) - \mu_{\tilde{S}_1}(\omega) \cdot \mu_{\tilde{S}_2}(\omega)\right)}, \quad \gamma \geqslant 0$$

$$t_{\mathrm{W}}^{\gamma}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) := \begin{cases} t_{\mathrm{D}}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) & \text{if } \gamma = -1 \\ t_{\mathrm{P}}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) & \text{if } \gamma = \infty \\ \max\left[\frac{\mu_{\tilde{S}_1}(\omega) + \mu_{\tilde{S}_2}(\omega) - 1 + \gamma \cdot \mu_{\tilde{S}_1}(\omega) \cdot \mu_{\tilde{S}_2}(\omega)}{1 + \gamma}, 0\right] & \text{otherwise} \end{cases}$$

$$t_{\mathrm{DO}}^{\gamma}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) := \frac{1}{1 + \left[\left(\frac{1}{\mu_{\tilde{S}_1}(\omega)} - 1\right)^{\gamma} + \left(\frac{1}{\mu_{\tilde{S}_2}(\omega)} - 1\right)^{\gamma}\right]^{1/\gamma}}, \quad \gamma > 0$$

$$t_{\mathrm{YY}}^{\gamma}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) := \max\left[(1 + \gamma)\left(\mu_{\tilde{S}_1}(\omega) + \mu_{\tilde{S}_2}(\omega) - 1\right) - \gamma \cdot \mu_{\tilde{S}_1}(\omega) \cdot \mu_{\tilde{S}_2}(\omega), 0\right], \quad \gamma \in \mathbb{R}$$

$$t_{\mathrm{Y}}^{\gamma}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) := 1 - \min\left[1, \left((1 - \mu_{\tilde{S}_1}(\omega))^{\gamma} + (1 - \mu_{\tilde{S}_2}(\omega))^{\gamma}\right)^{1/\gamma}\right], \quad \gamma > 1$$

The first of these is from Dubois and Prade (1982b), $t_{\mathrm{F}}^{\gamma}(\cdot, \cdot)$ comes from Frank (1979), $t_{\mathrm{H}}^{\gamma}(\cdot, \cdot)$ was proposed as a generalization of the Hamacher t-norm stated above and presented along with it in Hamacher (1978), $t_{\mathrm{W}}^{\gamma}(\cdot, \cdot)$ comes from Weber (1983), $t_{\mathrm{DO}}^{\gamma}(\cdot, \cdot)$ from Dombi (1982), $t_{\mathrm{YY}}^{\gamma}(\cdot, \cdot)$ from Yandong (1985), and $t_{\mathrm{Y}}^{\gamma}(\cdot, \cdot)$ from Yager (1980).

Finally, we look at two averaging operators:

$$\mathrm{A}_1^{\gamma}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) := \gamma \min\left[\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right] + (1 - \gamma)\frac{\mu_{\tilde{S}_1}(\omega) + \mu_{\tilde{S}_2}(\omega)}{2}, \quad 0 \leqslant \gamma \leqslant 1$$

$$\mathrm{A}_2^{\gamma}\left(\mu_{\tilde{S}_1}(\omega), \mu_{\tilde{S}_2}(\omega)\right) := \left(\mu_{\tilde{S}_1}(\omega) \cdot \mu_{\tilde{S}_2}(\omega)\right)^{1-\gamma}\left(\mu_{\tilde{S}_1}(\omega) + \mu_{\tilde{S}_2}(\omega) - \mu_{\tilde{S}_1}(\omega) \cdot \mu_{\tilde{S}_2}(\omega)\right)^{\gamma}, \quad 0 \leqslant \gamma \leqslant 1$$

The first averaging operator is from Luhandjula (1982), the second from Zimmermann and Zysno (1980, 1983). Both operators are so-called compensatory operators, meaning that a low degree of membership for one set can, to an extent determined by the parameter γ, be compensated by a high degree of membership for the other set. No t-norm has this feature. On the other hand, neither averaging operator is associative, and so they are not t-norms.

There is very limited previous empirical research on the above operators. Thole, Zimmermann, and Zysno (1979) looked at t_{M} and t_{P} and found these to do rather poorly. Of course, this does not preclude that, in contexts not considered in their research, these operators, or one of them, does well. Similarly, while Zimmermann and Zysno (1983) report positive results for A_2^{γ}, these results concern a specific decision-making context, and it remains to be seen how far the success of that averaging operator generalizes.

4.2 Method

Participants

Three hundred and thirty-nine participants from Australia, Canada, the United Kingdom, and the United States were recruited via CrowdFlower, where they were redirected to the Qualtrics platform on which the survey was run. Each participant was compensated with $2.50 (USD). Repeat participation was blocked via the Qualtrics software.

We excluded participants who returned incomplete response sets and participants who completed the survey in less than five minutes, which pilot testing had shown to be the minimally required time to read all the materials and questions. We further excluded nonnative speakers of English (given that the materials were in English), participants who indicated that they were color blind, participants who answered negatively to the question of whether they had responded seriously (adding this question to the survey followed a recommendation by Aust et al. 2014), and participants who made at least five mistakes on a color sorting task taken from Douven et al. (2017). This left 205 participants for the final analysis.

These participants spent on average 17.25 ($\pm$ 26.71) minutes on the survey. One hundred and five of them were female, 98 were male, and two preferred not to say. Their mean age was 36.65 ($\pm$ 11.51) years. Finally, 149 of these participants had a college degree, 43 indicated high school as their highest education level, and the remaining participants indicated a lower education level. A re-analysis of the data on the basis of all 339 of the original participants led to qualitatively identical results.

Materials and procedure

The materials for this study were put together from the materials used for the studies reported in Douven (2016b) and Douven et al. (2017), which were summarized in the previous section. Specifically, all shapes shown in Figure 6 were used, but now each shape appeared not in neutral gray but uniformly colored in one of the colors of the ten right-most patches of Series 1, shown in Figure 3. The number of colors used was limited to ten to make the survey not overly long while allowing each participant to judge a reasonably large proportion of the stimuli. The choice for the particular colors was based on the results from Douven et al. (2017). As seen in Figure 5, for the first four patches in Series 1, there is virtually no variability, whether we look at the predicted or at the observed degrees. Moreover, discarding those four colors yielded a well-balanced series, in that it contained equal numbers of clear and borderline cases as well as, among the clear cases, equal numbers of clear blue and clear green cases. The same balance could not have been obtained with either of the other series. The materials contained each shape from Figure 6 appearing in all ten colors, yielding a total of 490 stimuli.

Each participant was shown 60 out of these 490 stimuli, where the stimuli were selected randomly per participant and shown, one per screen, in a random order. Together with each stimulus appeared the question of how strongly the participant agreed that the image shown represented a blue vase. Participants were asked to answer this question by positioning a slider on a scale from -50 to 50. To these anchors were attached, next to the numbers, the labels "Strongly disagree" and "Strongly agree," respectively; the midpoint of the scale had the number 0 as well as the label "Neither agree nor disagree" attached to it. To avoid response bias as much as possible, the slider was initially placed at the midpoint. See Figure 9 for two examples of how the screen could look. At the beginning of the survey, participants were told that there were no "correct" responses and that whatever seemed right to them was a valid response.

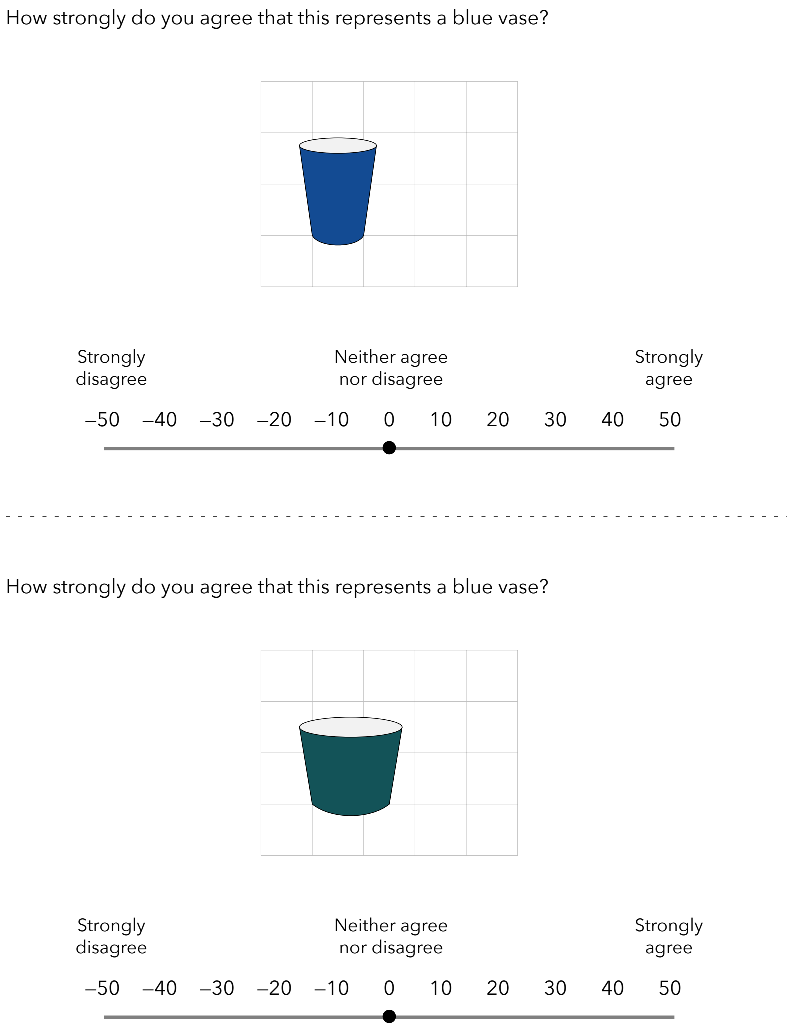

Figure 9: Two examples of screens that a participant could have seen during the survey.

4.3 Results and discussion

There were a total of $205 \times 490 = 100{,}450$ responses. Each stimulus was evaluated by 25.1 (± 3) participants, on average, with a minimum of 17 evaluations and a maximum of 32. The analysis sought to determine how well these responses could be predicted on the basis of the degrees of vasehood and degrees of blueness as determined separately in the empirical work discussed previously. It thereby aimed to answer research questions Q1–Q3.

To give an intuitive understanding of what the analysis was meant to accomplish, consider that, according to the predictions from Douven et al. (2017) shown in the top row of Figure 4, the shade of patch number 9 in Series 1 from Figure 3 has a degree of blueness of approximately .89 (assuming the largest prototypical regions; the degree assuming any of the other cutoff points is very close, however).[13] According to the predictions from Douven (2016b) shown in the top row of Figure 8, the degree of

[13]Both the predicted degrees of blueness of the used color shades and the predicted degrees of vasehood of the shapes of the stimuli are given numerically in the R file that is part of the Supplementary Materials.

vasehood of shape number 18 from Figure 6 has a degree of vasehood of approximately .78. We know from earlier research that these degrees accurately predict how strongly people agree that the said patch is blue, and that the said shape represents a vase, respectively. Now, is there a function $f(\cdot,\cdot)$ such that $f(.89,.78)$ accurately predicts how strongly people agree that the said shape appearing in the said color is a blue vase, and that similarly yields accurate predictions for our other stimuli? Or is there possibly a function $f(\cdot,\cdot,\gamma)$ that does this, for some value of γ? And finally, is there a non-parameterized operator $\otimes(\cdot,\cdot)$ or a parameterized one $\otimes_\gamma(\cdot,\cdot)$ among the operators listed above such that $f(\cdot,\cdot) = \otimes(\cdot,\cdot)$ or $f(\cdot,\cdot,\gamma) = \otimes_\gamma(\cdot,\cdot)$?

To answer these questions, we started by calculating, for each of the non-parameterized t-norms listed in Section 4.1 and for all 490 stimuli, degrees of blue vasehood on the basis of degrees of blueness (B) and degrees of vasehood (V) from previous research. We then fitted, for each of the resulting sets of degrees of blue vasehood, a linear mixed-effects model (LMM) with responses as the dependent variable and calculated degrees of blue vasehood as fixed effect; each model also included random intercepts and random slopes for participants (see Barr et al. 2013). The models were fit using the lme4 package (Bates et al. 2015) for R. For the parameterized t-norms and the averaging operators, the analysis proceeded along the same lines, except that here we calculated, for each of those operators and all stimuli, degrees of blue vasehood for a wide range of values of the parameter, and then fitted LMMs for each resulting set of degrees. (See the R script for details about the process of finding the best value for the parameter.) Finally, we fitted two additional LMMs, both with responses as dependent variable, and one with B and V as fixed effects and the other with B and V as well as their interaction as fixed effects; both models had the same random-effects structure as the previous ones. The relevant model comparison statistics are given in Table 1, where for the parameterized operators we report the best model.

A first observation to make is that model fit for the models with predictors based on non-parameterized t-norms was rather poor, with R^2-values ranging from .33 to .42. The parameterized t-norm models did not do any better, with the exception of the one based on an instance of the Weber t-norm schema, which achieved an R^2-value of .56. Of the parameterized averaging operators, the second one, which Zimmermann and Zysno (1983) had found to do well on their data, did best and reached acceptable fit—as indicated by an R^2-value of .64—for some values of the parameter. That some models based on a parameterized operator did much better than any of the models based on a non-parameterized t-norm is not a surprise, of course. It was mentioned that, according to Zimmermann, the best we can hope for is to have "a very small number of [aggregation] operators to model many situations," but in order to achieve that aim—Zimmermann (2001, p. 44) notes–these operators will have to be adaptable to whatever the application context is. And a key advantage of the parameterized over the non-parameterized operators is exactly that the former are adaptable, by allowing a per-context fine-tuning of the value of the parameter.

While an R^2-value of .64 is certainly acceptable, it is not an indication of *good* model fit. More importantly still, it shows that even the best of the models based on a known operator lags far behind the models that are *not* based on such an operator, and that use either B and V alone, or B and V as well as their interaction, as separate predictors. These models can boast of an unequivocally good model fit, with an R^2-value of .81 and .82, respectively. Because a model with a greater number of predictors can have better model fit just for that reason, we also compared the models using metrics that penalize for complexity, to wit, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which weigh model fit and model complexity against each other. The resulting statistics are to be used comparatively, in the sense that models with lower values are taken to be predictively more

Table 1: Comparison of the LMMs.

predictor(s)	k	LL	AIC	ΔAIC	BIC	ΔBIC	R^2	γ
$t_{\mathrm{M}}(\mathrm{B}, \mathrm{V})$	6	−58908.73	117829.5	13981.3	117874.0	13914.5	.41	–
$t_{\mathrm{P}}(\mathrm{B}, \mathrm{V})$	6	−58792.75	117597.5	13749.3	117642.0	13682.6	.42	–
$t_{\mathrm{L}}(\mathrm{B}, \mathrm{V})$	6	−58784.40	117580.8	13732.6	117625.3	13665.9	.42	–
$t_{\mathrm{D}}(\mathrm{B}, \mathrm{V})$	6	−59537.42	119086.8	15238.7	119131.3	15171.9	.33	–
$t_{\mathrm{H}}(\mathrm{B}, \mathrm{V})$	6	−58841.25	117694.5	13846.3	117739.0	13779.6	.41	–
$t_{\mathrm{E}}(\mathrm{B}, \mathrm{V})$	6	−58779.56	117571.1	13722.9	117615.6	13656.2	.42	–
$t^{\gamma}_{\mathrm{DP}}(\mathrm{B}, \mathrm{V})$	6	−58792.75	117597.5	14749.3	117642.0	13682.6	.42	1.00
$t^{\gamma}_{\mathrm{F}}(\mathrm{B}, \mathrm{V})$	6	−58773.98	117559.9	14711.7	117604.4	13645.0	.42	101.00
$t^{\gamma}_{\mathrm{H}}(\mathrm{B}, \mathrm{V})$	6	−58694.57	117401.1	14552.9	117445.6	13486.2	.43	79.00
$t^{\gamma}_{\mathrm{W}}(\mathrm{B}, \mathrm{V})$	6	−57206.20	114424.4	11576.2	114468.9	10509.5	.56	0.15
$t^{\gamma}_{\mathrm{DO}}(\mathrm{B}, \mathrm{V})$	6	−58643.82	117299.6	14451.4	117344.1	13384.7	.43	0.30
$t^{\gamma}_{\mathrm{YY}}(\mathrm{B}, \mathrm{V})$	6	−58416.39	116844.8	13996.6	116889.3	12929.9	.46	−9025.00
$t^{\gamma}_{\mathrm{Y}}(\mathrm{B}, \mathrm{V})$	6	−58553.02	117118.0	14269.8	117162.5	13203.1	.45	29290.00
$\mathrm{A}^{\gamma}_{1}(\mathrm{B}, \mathrm{V})$	6	−57220.77	114453.5	11605.3	114498.1	10538.7	.56	0.00
$\mathrm{A}^{\gamma}_{2}(\mathrm{B}, \mathrm{V})$	6	−55955.80	111923.6	9075.4	111968.1	8008.7	.64	0.89
B, V	10	−52209.16	104438.3	590.1	104512.5	553.1	.81	–
B, V, B $\times$ V	15	−51909.09	103848.2	0.0	103959.4	0.0	.82	–

Note: k is the number of parameters and LL the log-likelihood. AIC is the Akaike Information Criterion and BIC the Bayesian Information Criterion, two metrics that weigh model fit against model complexity. ΔAIC/ΔBIC gives the difference in AIC/BIC value between the given model and the model with the lowest AIC/BIC value. R^2 is the correlation between fitted and observed values. R^2-values were calculated using the r2 function in the sjstats package (Lüdecke 2017) for R; this function follows the recommendations of Nakagawa and Schielzeth (2013). For the models based on a parameterized t-norm or an averaging operator, the last column gives the value of γ which yielded the model with the lowest AIC value or, in cases where there was more than one such model, the model with the lowest R^2-value among the ones with the lowest AIC value.

accurate than ones with higher values. Burnham and Anderson (2002, p. 70 f) propose as a rule of thumb that a difference in AIC value greater than 10 is to be interpreted as indicating that the model with the higher value enjoys essentially no support from the data. Now consider that the difference in AIC value between the overall best model and the best of the models based on a known operator is over 9000! A look at the AIC values also reveals that, while the two best models both attain good fit, the second-best still receives no support from the data. That the best model is significantly better than the second-best, and so that including the interaction term is warranted, is confirmed by a ratio likelihood test for those models: $\chi(5) = 600.14, p < .0001$.

To see why the model with B, V, and their interaction as predictors does better (mostly *much* better) than the competition, we have a closer look at it. Table 2 states the regression results for this model, and Figures 10 and 11 plot the main effects and the interaction effects, respectively.

Clearly, degree of blueness had a much bigger impact on the responses than degree of vasehood. One possible reason for this is that the colors that participants saw were *real* while the vase shapes were all *stylized*. Perhaps even more importantly, there is a lot of variation in the shapes of vases that we encounter in our everyday lives, to the extent that some vases have shapes that we would regard as being very *dissimilar* to the shapes of our stimuli. The same does not hold for the other predictor: it is difficult to imagine a shade of blue that people might regard as being very dissimilar to the colors of the stimuli.

Table 2: Regression results for model with B, V, and B × V as predictors.

	b	SE	*β*	*t*	*p*
B	57.16	2.08	0.70	27.55	<.0001
V	2.04	0.67	0.02	3.03	.0028
B × V	12.39	1.37	0.12	9.07	<.0001

Note: Beta coefficients were obtained via the std_beta function in the R package sjstats.

Another observation to make is that the interaction is *highly* significant. What this means is that *how* one of B and V influences people's judgments of the degrees of vasehood of the stimuli differs significantly for different values of the other predictor. In general, it may be very hard to say anything definite about such interaction effects without engaging in the kind of empirical research reported here.

Although our results show that none of the known intersection operators did particularly well in predicting people's judgments of graded membership for our stimuli, they also suggest the definition of a *new* type of intersection operator, one that presumes the availability of relevant data (e.g., obtained in the way we obtained judgments of blue vasehood). Generally put, the proposal is to define

$$\mu_{\tilde{S}_1 \cap \tilde{S}_2}(\omega) \quad = \quad \frac{c_1 \mu_{\tilde{S}_1}(\omega) \; + \; c_2 \mu_{\tilde{S}_2}(\omega) \; + \; c_3 \big(\mu_{\tilde{S}_1}(\omega) \times \mu_{\tilde{S}_2}(\omega)\big)}{c_1 \; + \; c_2 \; + \; c_3},$$

where the coefficients c_i are to be estimated from whatever the data at hand are. (The division by the sum of the coefficients ensures that degrees of membership are in the right range.) Applying this idea to the case at issue, using the estimates from our best LMM, yields as an intersection operator the following instance of the above schema:

$$\mu_{\text{blue vase}}(\omega) \quad = \quad \frac{57.16\, \mu_{\text{blue}}(\omega) \; + \; 2.04\, \mu_{\text{vase}}(\omega) \; + \; 12.39\big(\mu_{\text{blue}}(\omega) \times \mu_{\text{vase}}(\omega)\big)}{71.59}.$$

Not surprisingly, applying this operator to B and V and then using the outcome as the predictor in a newly fit LMM yields a model whose fit is as good as that of the best model from our analysis.

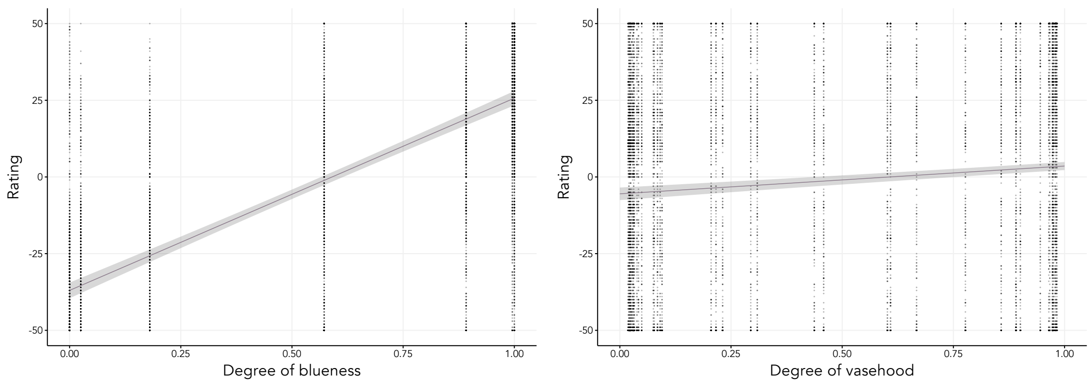

Figure 10: Plots of main effects as estimated by the best model. Data are plotted with jitter, to enhance visibility. Shaded bands indicate 95 percent confidence intervals.

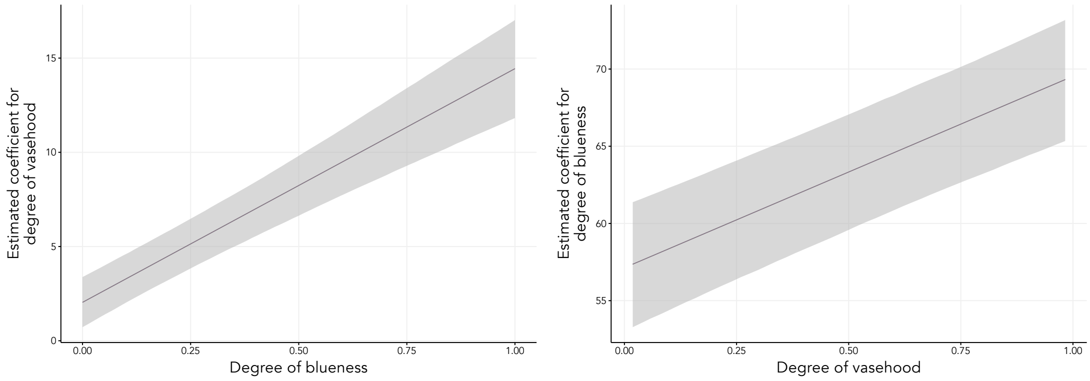

Figure 11: Plots of interaction effects as estimated by the best model. Shaded bands indicate 95 percent confidence intervals.

Note that the operator we here arrived at is not a t-norm; it need not even be a two-place function (given that the interaction between degree of blueness and degree of vasehood is not necessarily a function of those variables). Also note that, given this operator, the concept of blue vasehood is overextended, in the sense of Hampton (1987, 1988), meaning that it can apply to an item (to some degree) even if one of the constituent concepts does not apply to it. That may appear odd, but it is actually an effect that Hampton discovered to exist in his participants' responses. Rather than trying to predict graded membership judgments for conjunctive concepts on the basis of conceptual spaces, he tried to predict them on the basis of graded membership judgments for the constituent concepts. In line with our own results, he found that the former could be reliably predicted as a linear combination of the latter, together with an interaction term.

Some might object that defining an intersection operator on a case-by-case basis is an ad hoc manoeuvre. But why stick to the idea of having an intersection operator from those stated earlier (even if the parameterized ones allow for fine-tuning), when the chances of obtaining an operator that is, in Zimmermann's words cited previously, an "appropriate model of real-system behavior" are much better if we determine this operator on the basis of some collected data pertinent to the given context?[14] It is not that I do not see the theoretical advantage of having just one or two intersection operators cover all application cases. But that may be a luxury that we simply cannot afford, not, at least, if we want the operators to conform broadly to people's judgments. Those who still feel that my proposal strays too far from fuzzy set theory as it was originally meant to be should see the results reported in this paper as a challenge to come up with new intersection operators that do have the requisite features.[15]

[14] Bělohlávek and Klir (2011c, p. 59) already note that we may have to determine the most fitting intersection operator experimentally, and separately for each application context. However, for them this means choosing the most fitting operator from the extant ones, while our recommendation is to be open to the possibility that, at least sometimes, good fit can only be achieved by constructing an intersection operator in the way just explained.

[15] Zimmermann and Zysno (1980, p. 48) mention the possibility of generalizing their averaging operator A_2^γ by individually weighting the two sets involved such that the weights add up to 2. Obviously, one could generalize all of the operators from Section 4.1 in this way. But first, how would that put us in a better position than if we followed the procedure that I am recommending? There is certainly no sense in which it would be easier, or less costly, to estimate the parameters for such operators than it is to conduct the kind of regression analysis we conducted. Second, using a grid approximation procedure (see the R file), it was found that the best model for Zimmermann and Zysno's generalized averaging operator was obtained

As a more general objection, it might be said that, rather than showing something negative about any of the fuzzy intersection operators to be found in the previous literature, our results challenge the empirical adequacy of the definition of graded membership proposed in Decock and Douven (2014), the understanding being that in order for such a definition to be empirically adequate, it should not only make accurate predictions about atomic concepts but also about conjunctive concepts, where the intersection operator at stake must be one of those stated in Section 4.1. I fail to see, however, why fuzzy set theorists should be so strongly committed to those operators that they make the notion of empirical adequacy hostage to them. Indeed, the citations from Zimmermann (2001) given in the opening paragraphs of Section 4 suggest that at least some important proponents of fuzzy set theory have a rather open-minded and flexible attitude vis-à-vis the choice of intersection operator. Be that as it may, it is worth reiterating a point made in the same section, viz., that presently the operational definition of graded membership from Decock and Douven (2014) is our best hope of countering what otherwise appears a fundamental criticism of fuzzy set theory: that it builds off a mysterious notion of fuzzy set inclusion. So those who want to stick, at any cost, to one or a few of the intersection operators from Section 4.1 owe us an alternative reply to that criticism.

Finally, one may wonder how general our finding is. While the two conceptual spaces we relied on are not in any way exotic, it could still be that, for most conjunctive concepts, we can get by with a small subset of the known intersection operators. That there might then still be a couple of anomalous cases, the one we considered among them, would not be a disaster (even if also not ideal). We noted in Section 3 that if one wants to *work* with conceptual spaces, then there is only a very limited choice at the moment. The spaces used for the present study are, to the best of my knowledge, the only two that are readily available to researchers. Needless to say, that is an unfortunate situation for reasons that go well beyond the purposes of this paper. For now, all I can say is that it would be hugely surprising if the fuzzy intersection operators that were shown to fail for the families of concepts representable in the two currently available conceptual spaces would work just fine for virtually all other families of concepts. Admittedly, though, only further empirical work, notably including work to ensure the availability of a wider range of conceptual spaces, will be able to settle the question of the generality of our finding.[16]

5 General discussion

Previous empirical work supported the validity of an operational definition of graded membership that had been proposed in response to concerns about the scientific status of fuzzy set theory. However, that work provided data only on atomic fuzzy concepts, and thus did not address any concerns one might have about the various operations on fuzzy sets proposed by Zadeh and others. The present paper focused on one particular such operation, to wit, that of fuzzy intersection. The first question we asked, Q1, was whether the same definition of graded membership would also prove valid if applied to the intersection of fuzzy sets. A follow-up question, Q2, which presupposed Q1's having a positive answer, was whether the available information about the atomic concepts would be enough to accurately predict judgments concerning the degrees to which items are in their intersection. A second follow-up question, Q3, which in turn presupposed Q2's having a positive answer, was

by setting $\gamma = 0.956$ and assigning blueness and vasehood weights of 1.63 and 0.37, respectively. That model still did worse, across all relevant criteria, than our LMM with predictors B and V as well as their interaction; in particular, for the former model AIC = 105091.4 (versus 103848.2 for our best model), BIC = 105135.9 (versus 103959.4), and $R^2 = .80$ (versus .82).

whether such accurate predictions could be derived from the information about the atomic concepts by one of the intersection operators that have been proposed in the literature.

To answer these questions, we conducted a study with materials based on those of the empirical work on atomic concepts. The results showed that the definition of graded membership that had earlier yielded such accurate predictions for degrees of blueness and degrees of vasehood also yielded accurate predictions for degrees of blue vasehood, thereby answering Q1 in the positive. Note that there was nothing in previous empirical work to suggest this answer. Decock and Douven's (2014) definition of graded membership could have worked for atomic concepts but then have failed for anything going beyond those.

Furthermore, obtaining a *yes* to Q1 could have required constructing a space of blue vasehood, say, by letting participants judge the similarity of pairs of our new stimuli and then using the resulting judgments as input data for an MDS procedure. But this turned out to be unnecessary: we were able to accurately predict people's judgments of degrees of blue vasehood by relying strictly on our knowledge of the degrees of blueness of the shades of our stimuli and the degrees of vasehood of the shapes of those stimuli. Thus, Q2 has a positive answer as well.

At the same time, we saw that using any of the known intersection operators to calculate degrees of blue vasehood from degrees of blueness and degrees of vasehood led to mostly disappointing results. No LMM that tried to predict participants' responses on the basis of degrees calculated in any of the said ways reached good model fit. Hence, the answer to Q3 must be negative.

We mentioned that Zimmermann emphasized the importance of an intersection operator's being *adaptable*. His point was underpinned by the finding that some parameterized operators did manifestly better on our data than all of the non-parameterized ones. But it was seen that, even of the former, none did really well. Indeed, one important lesson from our experimental results is that, in general at least, a single adjustable parameter is not enough for an intersection operator. It is not enough to be able to simply weigh differently the sets that are being intersected. Our best LMM, which had very good model fit, included an interaction term that was highly significant. From this model, we could read off a three-parameter operator. To be sure, in some contexts fewer parameters, or even a non-parameterized operator, may do. Generally speaking, however, to what extent the degrees of membership for the individual concepts impact on the degree of membership for the conjunctive concept, and to what extent the interaction between the former contributes to the latter, will most likely have to be considered separately per application context. Admittedly, this gives fuzzy set theory somewhat of an open-ended character, and makes it more of a paradigm than a theory *stricto sensu*. But that is an idea which some in the fuzzy set theory community of researchers are already sympathetic to on largely independent grounds (e.g., Bělohlávek & Klir 2011b, 2011c).[17]

References

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods* 45:372–400.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.

[17] Thanks to two anonymous referees for this journal as well as to Christopher von Bülow for valuable comments on previous versions. I am also grateful to an audience at the University of Louvain-la-Neuve for stimulating questions and discussion.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). lme4: *Linear mixed-effects models using eigen and S4*. R package version 1.1–0 (available at http://lme4.r-forge.r-project.org/).
Bělohlávek, R., Dauben, J. W., & Klir, G. J. (2017). *Fuzzy logic and mathematics: A historical perspective*. New York: Oxford University Press.
Bělohlávek, R., & Klir, G. J. (2011a). Epilogue. In R. Bělohlávek & G. J. Klir (eds.), *Concepts and fuzzy logic*. Cambridge MA: MIT Press, pp. 259–264.
Bělohlávek, R., & Klir, G. J. (2011b). Fallacious perceptions of fuzzy logic in the psychology of concepts. In R. Bělohlávek & G. J. Klir (eds.), *Concepts and fuzzy logic*. Cambridge MA: MIT Press, pp. 121–148.
Bělohlávek, R., & Klir, G. J. (2011c). Fuzzy logic: A tutorial. In R. Bělohlávek & G. J. Klir (eds.), *Concepts and fuzzy logic*. Cambridge MA: MIT Press, pp. 45–87.
Berlin, B., & Kay, P. (1969). *Basic color terms*. Stanford CA: CSLI Publications.
Borg, I., & Groenen, P. (2010). *Modern multidimensional scaling* (2nd ed.). New York: Springer.
Borgwardt, S., Cerami, M., & Peñaloza, R. (2017). The complexity of fuzzy $\mathcal{EL}$ under the Łukasiewicz T-norm. *International Journal of Approximate Reasoning* 91:179–201.
Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: A practical information-theoretic approach*. Berlin: Springer.
Decock, L., & Douven, I. (2014). What is graded membership? *Noûs* 48:653–682.
Dombi, J. (1982). A general class of fuzzy operators, the De Morgan class of fuzzy operators and fuzziness induced by fuzzy operators. *Fuzzy Sets and Systems* 8:149–163.
Douven, I. (2016a). *The epistemology of indicative conditionals*. Cambridge: Cambridge University Press.
Douven, I. (2016b). Vagueness, graded membership, and conceptual spaces. *Cognition* 151:80–95.
Douven, I. (2018). A Bayesian perspective on Likert scales and central tendency. *Psychonomic Bulletin and Review* 25:1203–1211.
Douven, I. (2019a). The rationality of vagueness. In R. Dietz (ed.), *Vagueness and rationality*. New York: Springer, pp. 115-134.
Douven, I. (2019b). New foundations for fuzzy set theory. In A. Aberdein & M. Inglis (eds.), *Advances in experimental philosophy of logic and mathematics*. London: Bloomsbury, pp. 173–199.
Douven, I. (2019c). Putting prototypes in place. *Cognition* 193:104007, https://doi.org/10.1016/j.cognition.2019.104007.
Douven, I., & Decock, L. (2017). What verities may be. *Mind* 126:386–428.
Douven, I., Decock, L., Dietz, R., & Égré, P. (2013). Vagueness: A conceptual spaces approach. *Journal of Philosophical Logic* 42:137–160.
Douven, I., & Gärdenfors, P. (2020). What are natural concepts? A design perspective. *Mind & Language*, in press.
Douven, I., Wenmackers, S., Jraissati, Y., & Decock, L. (2017). Measuring graded membership: The case of color. *Cognitive Science* 41:686–722.
Dubois, D., & Prade, H. (1979). Fuzzy real algebra: Some results. *Fuzzy Sets and Systems* 2:327–348.
Dubois, D., & Prade, H. (1980). *Fuzzy sets and systems: Theory and applications*. Boston: Academic Press.
Dubois, D., & Prade, H. (1982a). Towards fuzzy differential calculus. *Fuzzy Sets and Systems* 8:1–17; 105–116; 225–233.
Dubois, D., & Prade, H. (1982b). A class of fuzzy measures based on triangular norms. *International*

Égré, P., & Barberousse, A. (2014). Borel on the heap. *Erkenntnis* 79:1043–1079.

Fairchild, M. D. (2013). *Color appearance models*. Hoboken NJ: Wiley.

Fermüller, C. G. (2015). Semantic games for fuzzy logics. In P. Cintula, C. G. Fermüller, & C. Noguera (eds.), *Handbook of mathematical fuzzy logic* (vol. 3). London: College Publications, pp. 969–1028.

Frank, M. J. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y = F(x, y)$. *Aequationes Mathematicae* 19:194–226.

Gärdenfors, P. (2000). *Conceptual spaces*. Cambridge MA: MIT Press.

Gärdenfors, P. (2014). *The geometry of meaning*. Cambridge MA: MIT Press.

Haack, S. (1979). Do we need "fuzzy logic"? *International Journal of Man–Machine Studies* 11:437–445.

Hamacher, H. (1978). *Über logische Aggregationen nicht-binär expliziter Entscheidungskriterien*. Frankfurt/Main: Rita G. Fischer Verlag.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition* 15:55–71.

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* 14:12–32.

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science* 31:355–384.

Jraissati, Y., & Douven, I. (2018). Delving deeper into color space. *i-Perception* 9:1–27, https://doi.org/10.1177/2041669518792062.

Kamp, H. (1975). Two theories of adjectives. In E. L. Keenan (ed.), *Formal semantics of natural language*. Cambridge: Cambridge University Press, pp. 123–155

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

Keefe, R. (2000). *Theories of vagueness*. Cambridge: Cambridge University Press.

Klir, G. J., & Folger, T. A. (1988). *Fuzzy sets, uncertainty and information*. Englewood Cliffs NJ: Prentice-Hall.

Klir, G. J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications*. Upper Saddle River NJ: Prentice-Hall.

Levinson, S. C. (2000). *Presumptive meanings*. Cambridge MA: MIT Press.

Lindley, D. V. (2004). Comment. *Journal of the American Statistical Association* 99:877–879.

Lüdecke, D. (2017). sjstats*: Statistical functions for regression models*. R package version 0.8.0 (available at https://CRAN.R-project.org/package=sjstats).

Luhandjula, M. K. (1982). Compensatory operators in fuzzy linear programming with multiple objectives. *Fuzzy Sets and Systems* 8:245–252.

Malacara, D. (2002). *Color vision and colorimetry: Theory and applications*. Bellingham WA: SPIE Press.

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

Mizumoto, M., & Tanaka, K. (1982). Fuzzy sets and their operations. *Information and Control* 48:30–48.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4:133–142.

Okabe, A., Boots., B., Sugihara, K., & Chiu, S. N. (2000). *Spatial tessellations* (2nd ed.). New York: Wiley.

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition* 9:259–288.

Paris, J. B. (2000). Semantics for fuzzy logic supporting truth functionality. In V. Novák & Irina Perfilieva (eds.), *Discovering the world with fuzzy logic*. Heidelberg: Springer, pp. 82–104.

Rosch, E. (1973). Natural categories. *Cognitive Psychology* 4:328–350.

Rosch, E. (2011). "Slow lettuce": Categories, concepts, fuzzy sets, and logical deduction. In R. Bělohlávek & G. J. Klir (eds.), *Concepts and fuzzy logic*. Cambridge MA: MIT Press, pp. 89–120.

Russell, B. (1923). Vagueness. *Australasian Journal of Philosophy and Psychology* 1:84–92. (Reprinted in R. Keefe & P. Smith (eds.), *Vagueness: A reader*. Cambridge MA: MIT Press, 1997, pp. 61–68. The page reference is to the reprint.)

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:1411–1436.

Thole, U., Zimmermann, H.-J., & Zysno, P. (1979). On the suitability of minimum and product operators for the intersection of fuzzy sets. *Fuzzy Sets and Systems* 2:167–180.

Verheyen, S., & Égré, P. (2018). Typicality and graded membership in dimensional adjectives. *Cognitive Science* 42:2250–2286.

Weber, S. (1983). A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms. *Fuzzy Sets and Systems* 11:115–134.

Williamson, T. (1994). *Vagueness*. London: Routledge.

Wright, C. (1992). Is higher order vagueness coherent? *Analysis* 53:193–200.

Yager, R. R. (1980). On a general class of fuzzy operators. *Fuzzy Sets and Systems* 4:235–242.

Yandong, Y. (1985). Triangular norms and TNF-sigma algebras. *Fuzzy Sets and Systems* 16:251–264.

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

Zadeh, L. (1975). Fuzzy logic and approximate reasoning. *Synthese* 30:407–425.

Zardini, E. (2013). Higher-order sorites paradox. *Journal of Philosophical Logic* 42:25–48.

Zimmermann, H.-J. (2001). *Fuzzy set theory—and its applications* (4th ed.). New York: Springer.

Zimmermann, H.-J., & Zysno, P. (1980). Latent connectives in human decision making. *Fuzzy Sets and Systems* 4:37–51.

Zimmermann, H.-J., & Zysno, P. (1983). Decisions and evaluations by hierarchical aggregation of