Vilnius University Press

Contents lists available at Vilnius University Press

Problemos
2024 Priedas, pp. 58–77

ISSN 1392-1126 eISSN 2424-6158
DOI: https://doi.org/10.15388/Problemos.Priedas.24.5

An Empirical Comparison of Semantics for Quantified Vague Sentences

Alexandre Cremers

Faculty of Philology
Vilnius University
Email alexandre.cremers@gmail.com
ORCID 0000-0002-7435-6515

Julija Kalvelyte

Faculty of Philology
Vilnius University
Email julija.kalvelyte@gmail.com

Abstract. We investigate the compositional semantics of vague quantified sentences, focusing on sentences such as "All of the students are tall," where a non-vague quantifier quantifies into a vague predicate. While much work has been done on vagueness in natural language, including the semantics of vague adjectives, little attention has been paid so far to how vagueness interacts with complex sentences. We present an experiment that gathers data on naïve speakers' interpretation of such sentences after collecting their judgment on the applicability of the vague predicate for each individual in the restrictor. We then compare how three prominent fuzzy logics – Gödel, product, and Łukasiewicz – predict the acceptability of the quantified sentences. Our results indicate that Gödel logic best matches human behavior. We then prove an equivalence between Gödel logic and a probabilistic form of Williamson's epistemicism for the sentences we have tested, and discuss how our findings inform the broader debate on the semantics of vagueness, particularly between epistemicism and graded-truth approaches.
Keywords: vagueness, fuzzy logic, semantics, psycholinguistics

Empirinis kvantifikuotų neapibrėžtų sakinių semantikos palyginimas

Santrauka. Nagrinėjame neapibrėžtų kvantifikuotų sakinių kompozicinę semantiką, daugiausia dėmesio skirdami tokiems sakiniams kaip „Visi studentai aukšti", kuriuose apibrėžtas kvantorius kvantifikuoja neapibrėžtą predikatą. Nors natūraliojoje kalboje neapibrėžtumas jau yra plačiai tyrinėtas, įskaitant ir neapibrėžtų būdvardžių semantiką, lig šiol mažai dėmesio skirta neapibrėžtumo sąveikai su sudėtiniais sakiniais. Pristatome eksperimentą, kurio metu surinkome duomenis, kaip naivūs kalbėtojai interpretuoja tokius sakinius. Surinkę jų sprendimus apie neapibrėžto predikato taikytinumą kiekvienam ribotinio sakinio individui, lyginame, kaip trys žinomos neraiškiosios logikos – Gödelio, produkto ir Łukasiewicziaus – prognozuoja kvantifikuotų sakinių priimtinumą. Gauti rezultatai atskleidė, kad Gödelio logika tiksliausiai atitinka dalyvių supratimą. Galiausiai įrodome, kad teste pasirodžiusių sakinių atžvilgiu Gödelio logika ir tikimybinė Williamsono epistemizmo forma yra ekvivalenčios, ir aptariame, kaip mūsų išvados prisideda prie platesnės diskusijos apie neapibrėžtumo semantiką, ypač tarp epistemizmo ir laipsniuoto teisingumo požiūrių.
Pagrindiniai žodžiai: neapibrėžtumas, neraiški logika, semantika, psicholingvistika

Acknowledgement. We are thankful to Paul Egré, James Hampton, Timothy Williamson, and the audience of the 2nd Logic and Philosophy Conference for thoughtful discussions.
This research was made possible by funding from the Research Council of Lithuania and the European Social Fund under Measure 09.3.3-LMT-K-712.

Received: 06/10/2024. **Accepted:** 20/11/2024

1. Introduction

Language, when used by humans, is very often vague (Russell 1923). Even under modern definitions of vagueness, which are more restrictive than Russell's, one would be hard-pressed to hold a casual conversation without uttering a single vague term. A commonly accepted characterization of vagueness nowadays involves three key properties, which we will soon explain in detail: (a) the existence of borderline cases, (b) higher-order vagueness, and (c) the possibility to build Sorites paradoxes.

First, borderline cases designate objects which seem to fall neither in the positive nor negative extension of a predicate. For instance, a person of an average height would qualify as neither 'tall' nor 'not tall'. Higher-order vagueness builds on this notion, by stating that being a borderline case is itself a vague property: we cannot properly delineate between 'borderline tall' and 'clearly tall'. Crucially, this property extends recursively: 'borderline borderline tall' and 'borderline clearly tall' are again vague, and so on. Finally, vague terms allow Sorites paradoxes. The Sorites paradox dates back to the 4th century BCE and is attributed to Eubulides of Miletus, and can be formulated as such:

(1) A single grain of sand does not form a heap
If something is not a heap, adding a single grain of sand will not make it a heap
---
No amount of sand constitutes a heap

The paradox lies in the fact that both premises feel intuitively true, the inference seems valid, and yet the conclusion feels clearly false.

Numerous approaches have been tried to propose a semantics for vagueness in natural language which would account for all its properties. These include trivalent logics, which introduce a third truth value beyond 'true' and 'false', as first introduced by Hallden (1949) for other phenomena and applied specifically to vagueness by Tye (1994); for instance, fuzzy logic, developed by Zadeh (1975), which allows for continuous degrees of truth in the interval [0, 1] (see also Lakoff 1973); and supervaluationism, which handles vagueness by considering all the different ways of making a vague term precise at once (Fine 1975). Some theories also posit multiple notions of truth, such as Tolerant-Classical-Strict semantics (Cobreros et al. 2012), while others, like Williamson's (1994) epistemic approach, seek to maintain classical semantics by treating vagueness as a form of ignorance on the actual semantic denotation. Probabilistic approaches have gained in popularity recently (Lassiter & Goodman 2014, 2017; Qing & Franke 2014; van Tiel et al. 2021; Xiang et al. 2022; Cremers 2022a). Most of them can be seen as implementations of the epistemic approach (the semantics is binary, but introduces free variables, the values of which can only be inferred approximately via pragmatic reasoning). Some even take into account multiple sources of uncertainty (within and between speakers, over time, across paradigms; see Cremers 2022b, 2024; Sarafoglou et al. 2024). A notable exception would be Carcassi et al.'s (2021) work on 'most', which assumes a fully-determined precise semantics and attributes vagueness exclusively to pragmatic uncertainty about the communicative intentions of the speaker, but this view would not extend to other vague terms such as relative gradable adjectives. Simultaneously, graded-truth approaches have also seen a resurgence of interest following Douven (2016).

With so many options, a key challenge is to effectively evaluate and compare these diverse semantic accounts of vagueness. Philosophers have constructed careful arguments to refute some of these proposals (based, for instance, on different predictions for contradictions and tautologies), but these arguments ultimately rely on introspective judgments. While introspective judgments on linguistic matters are surprisingly robust (Sprouse et al. 2013; Marty et al. 2020), vagueness tends to muddy the water. The field has therefore long relied on quantitative studies to clarify important empirical questions (e.g., Rosch 1973; Hersh & Caramazza 1976).

2. Compositional Semantics for Vague Quantified Sentences

Much fewer accounts have looked at how vagueness interacts with compositional semantics, and to the best extent of our knowledge, none makes quantitative predictions. There are, of course, discussions of gradable adjectives in compositional semantics (e.g., Kennedy 1999; Kennedy & McNally 2005), but vagueness is rarely discussed in the context of complex sentences, and, even if it is, it is only performed in binary terms (a sentence is predicted to either show vagueness or not). For instance, a question that bothered semanticists for some time is how constructions such as "6 feet tall" or "taller than Mary" can be non-vague when the bare adjective 'tall' is vague. While empirical work on simple connectives suggests that fuzzy logic offers a good approximation of conjunction, negation, and possibly disjunction (Hersh & Caramazza 1976; Leffel et al. 2019, a.o.), Douven (2021) shows that no fuzzy conjunction offers a good model of combination (e.g., to what degree an object counts as a 'blue vase' given how 'blue' and how 'vase' it is).

Of the proposals discussed in the introduction, some would more easily generalize to complex sentences. This is particularly true of the various flavors of fuzzy logic, for which first-order logics have been proposed, but also for all work stemming from the epistemic approach to vagueness. Indeed, the epistemic approach states that semantics is classical, and that the thresholds for vague terms are fixed, but unknown. This translates directly into a semantics where composition remains binary, but the thresholds project as free variables. In any given situation, we can then use our probabilistic beliefs on the distribution of the thresholds to calculate a probability of the whole sentence being true.

Supervaluationism also extends immediately to complex sentences: the semantics for each precisification is classical, and the super-evaluation can take place at the level of the whole sentence. Yet, this only makes categorical predictions (a sentence is either super-true, super-false, or neither). Probabilistic versions of supervaluationism have been proposed recently in Spector (2017) and Cremers (2022b), but the former does not discuss vagueness, whereas the latter only adopts supervaluationism at the pragmatic level as a way to deal with second-degree vagueness: it presupposes that precisifications themselves are probabilistic or graded, making it closer to the 'plurivaluationalist fuzzy' approach (Smith 2011) or fuzzy epistemicism (MacFarlane 2010).

3. Experiment

3.1. Goal

Given how little is known about the meaning of quantified vague sentences, our first and main goal is to provide a quantitative description of naïve speakers' interpretation of these sentences. This data could then be used to build a descriptive model of the compositional semantics, with potential applications, especially in pragmatics (see Cremers 2022b, for a tentative pragmatics of vague sentences).

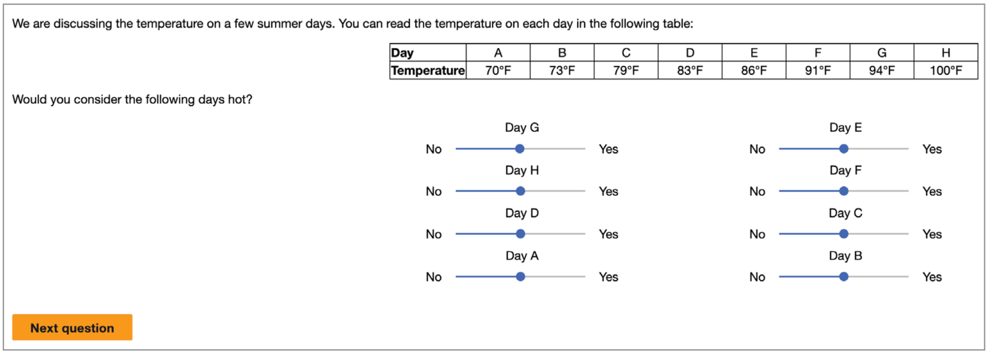

Figure 1. Example screen for the first step (measuring the applicability of the vague term to each individual item)

A secondary goal will be to evaluate the various proposed models for the semantics of vagueness. Note, however, that we can only evaluate models which make quantitative predictions for the kind of quantified sentences we will study. Some models do not make any quantitative predictions, while some need to be complemented with additional assumptions to cover the relevant fragment. This means that our results will be most useful when comparing closely related models, which share enough assumptions. In particularly, we will show how they can be used to distinguish between different flavors of fuzzy logic. In the discussion, we will explain the broader impact of our results for theories of vagueness other than graded-truth.

We will focus on a rather restricted set of quantified sentences, namely, sentences of the form $D(A)(B)$ where D is a non-vague determiner, and A, B are two predicates on a set of items E such that B but not A is vague in the given context (so that $D(A)$ is a non-vague generalized quantifier). We will only consider cases where all elements in the restrictor can be assumed to belong to the same comparison class. This is in contrast with examples such as (2) from Kennedy (1999), where each person is compared to a possibly different comparison class.

(2) Everyone in my family is tall
$\rightsquigarrow$ every member of the family is tall compared to people of their age, gender. . .

None of the quantifiers used in the experiment are vague (the only source of vagueness is the predicate in the scope of the quantifier). While there has been ample work on vague quantifiers such as ‘few’, ‘many’, or ‘most’ (Glöckner 2008; Pietroski et al. 2009; Kotek et al. 2015; Solt 2016; Schöller & Franke 2017; van Tiel et al. 2021; Denić & Szymanik 2022; Ramotowska et al. 2024; Sarafoglou et al. 2024, a.o.), testing their interaction with vague predicates would further complicate the matter, and we leave this for future research.

3.2. Methods and Materials

To probe the compositional semantics, the design we adopted consisted of two parts.

First, we measured the applicability of a vague predicate *B* to every element in *E*. For this purpose, we provided measurements for each element *x* in *E* along a scale relevant to *B* (e.g., height for ‘tall’). Our sets always contained 8 elements, and the applicability of *B* was measured for all elements at once, as shown in Figure 1. The applicability was measured by using sliders, and the order of the sliders for the 8 elements was randomized. After adjusting all sliders, the participants could click on a *Next question* button to move to the next part of the experiment.

Second, we measured the acceptability of $D(A)(B)$ for various D and A (four per participant), all using the same set *E* and the same vague predicate *B*. The participants were told explicitly which elements of *E* belonged to *A* and which did not, so that even if some *A* predicates could potentially accept some borderline cases, they would behave classically for the purpose of the experiment[1]. The four examples each participant saw were presented sequentially: between each example, the participants had to click a *Next question* button. An example is provided in Figure 2.

The rationale behind this design is that we can ultimately eliminate the provided measurements for elements of *E* and directly relate the acceptability of $D(A)(B)$ to the provided (binary) $A(x)$ and the measured (graded) $B(x)$ for each x in *E*.

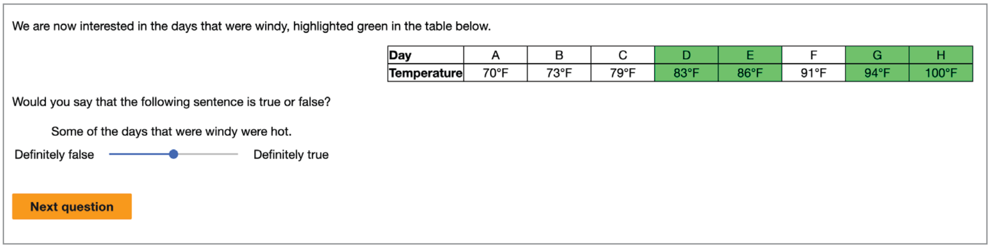

Figure 2. Example screen for the second step (measuring the acceptability of a quantified vague sentence)

Each participant saw only one gradable adjective for the quantifier scope *B* taken from a list of 12, as it would have been impractical to repeat Step 1 for multiple adjectives.

[1] Remember that no measurement was given for any dimension supporting *A*, so the participants only received binary information on which items satisfied *A* and which did not.

Each vague adjective was associated with a specific set of items and measurement along a corresponding dimension, as well as four restrictors *A* that were used to build the four quantified sentences the participant would judge. The stimuli were adapted from Cremers (2024).

Semanticists have identified several types of gradable adjectives: relative adjectives, like ‘tall’, are highly context-dependent and vague. By contrast, absolute adjectives like ‘full’ have a clear boundary to their scale, and, while they tolerate some approximation, they are usually not considered vague because the speaker could in principle agree to place the threshold at the known boundary (see, e.g., Kennedy & McNally 2005; McNally 2011, but also Burnett 2014, for counterpoints). Among absolute adjectives, semanticists further distinguish between *minimum* and *maximum* adjectives. As the names suggest, the difference is whether the relevant boundary is at the bottom or at the top end of the underlying scale. The terms *partial* and *total* are also sometimes used in the literature for this distinction.

The four quantified sentences the participants saw each involved a different quantifier, taken from a list of 7: ‘two’, ‘three’, ‘some’, ‘all’, ‘none’, and conjunctions of 2 or 3 items. For conjunctions, the ‘restrictor’ simply corresponded to the two or three items mentioned in the conjunction (so the actual predicate did not appear in the target sentence).

By varying the items that matched the restricting predicate, we could manipulate the intended type of the case under consideration (true, false, or borderline). Several important clarifications are in order here. First, there is no theory-neutral way to predict which situations will correspond to a ‘borderline case’ for a complex vague sentence. Our classification only reflects our best guesses. Second, in some cases, we were limited by practical constraints. For instance, to design the ‘clearly true’ case for a triple conjunction, the best we can do is to select the top three of eight items. Yet, some participants might not consider the third highest item to be clearly in the extension of the vague predicate, and may, therefore, not give the highest rating to our intended ‘clearly true’ case. The third and most important point, however, is that we fortunately do not need to worry too much about this. Just like the measurements we have provided to elicit applicability judgments at Step 1, the categories we assign to the different restrictors at Step 2 are only used to make sure that each participant is exposed to diverse cases, from clearly false to clearly true. The categories play no direct role in the analysis of the results[2]. With all this in mind, for each quantifier, we designed four combinations meant to represent a clear true case, a clear false case, and two kinds of borderline cases. The concrete specifications of each case are given in Table 2. The restrictors for the four sentences each participant saw were taken to include one of each case. Remember, however, that each sentence involved a different quantifier, so we had to balance across the participants which quantifier is associated with which case.

2 The only role is that we will use the ‘clear true’ and ‘clear false’ cases to detect the participants who clearly did not take the task seriously, and we will evaluate the models on ‘borderline cases’ only, since all models make very similar predictions for clear true and false cases.

In all cases, we had to balance and randomize the associations between the quantifiers, cases, adjectives, and predicate used to implement each restrictor. Since it was impossible to test all combinations[3], we simply generated these associations randomly and checked after drawing that we had not inadvertently introduced significant imbalance or associations; this was achieved by using χ^2-tests.

3.3. Participants

We recruited 295 participants on *Amazon's Mechanical Turk* (we had aimed for 300, but some participants did not complete the task). The participants were compensated $0.57 for their participation, and the survey took them about 4 minutes. We excluded ten participants who gave cardinally opposite responses to the True and False cases they received (i.e., by placing the slider above 50% in the *False* case, or below 50% in the *True* case), and three more participants who left more than half of the sliders on their default position without adjusting them.

3.4. Results

[All data and scripts are available at https://github.com/Alex-Cremers/semantics-quantified-vague].

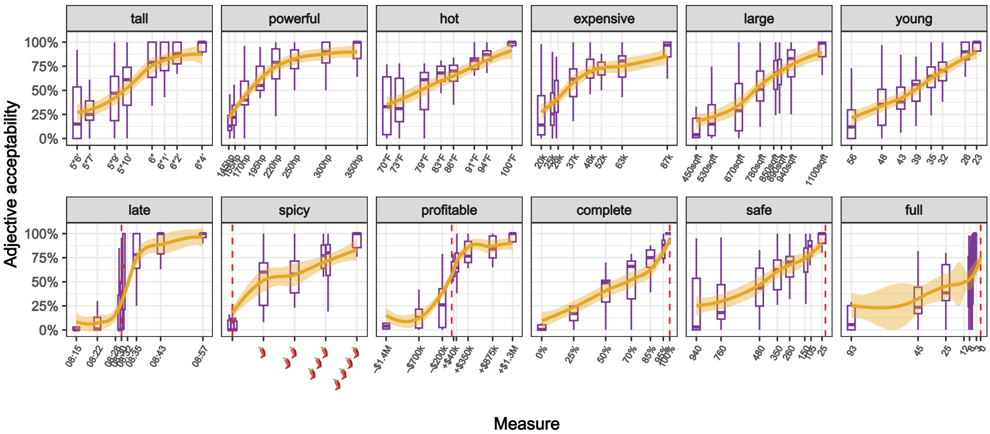

Figure 3. Rating on all items in the comparison class for each adjective. The purple boxplots indicate the distribution of the participants' ratings at each point in the scale, and an orange LOESS trend curve is superimposed. The top six adjectives are relative and lack a natural threshold. The bottom six are the so-called absolute adjectives, and can – in principle – receive a strict denotation at the threshold marked by a dashed vertical red bar

3 There are $\binom{7}{4} = 35$ ways to select 4 quantifiers out of 7, $4! = 24$ ways to assign the four truth cases to the selected four quantifiers, again 24 ways to assign a unique restrictor predicate from the rightmost column of Table 1 to each quantifier, and 12 adjectives to select from, resulting in 241,920 unique combinations, before we even consider the many possible orders of presentation.

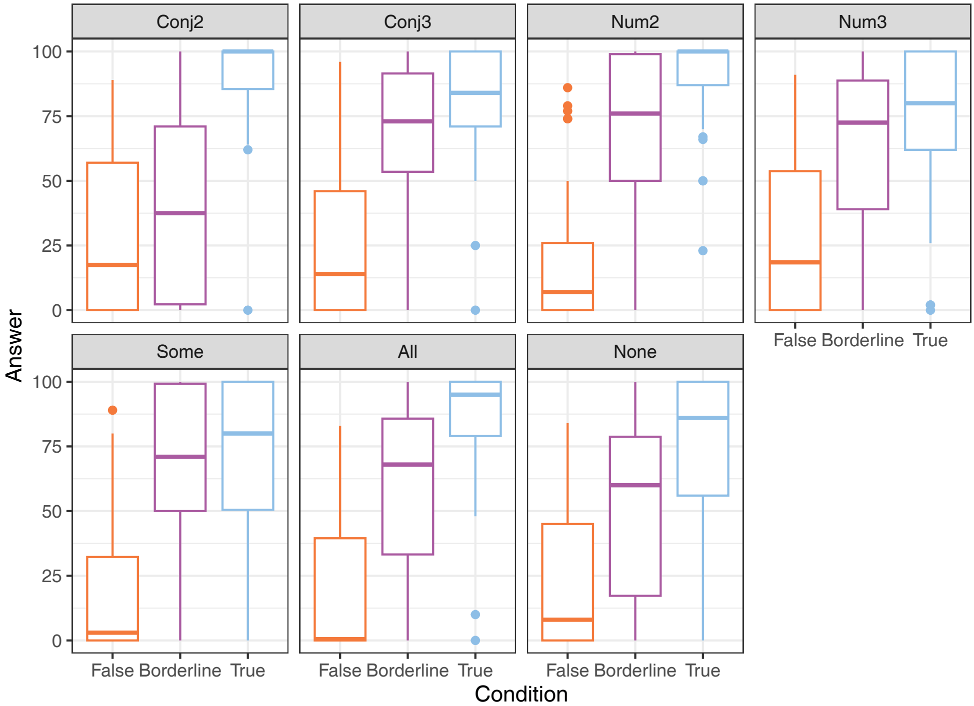

Figure 4. Distribution of the participants' judgments on each type of quantified sentence, by condition

Figure 3 shows the applicability ratings collected in the first question. As expected, the relative adjectives display vagueness in that there is no clear threshold from which all participants would agree that the adjective becomes applicable. For absolute adjectives, the applicability still does not follow a simple step function at the theoretical threshold, thereby suggesting that they are also subject to some uncertainty or imprecision[4]. Whether this is genuine vagueness or a separate phenomenon is debated in the semantics literature, as discussed above.

Figure 4 shows the distribution of the participants' judgments by condition and quantifier. Remember that the assigned label (True/Borderline/False) comes from the experimental design and should not be interpreted as a fact regarding the actual status of these sentences, nor as an embrace of trivalence on our part. Different proposals might predict a different status to the same sentence in the same condition. Importantly for us, on sentences assigned to the Borderline condition – but also to a lesser extent on those assigned True or False – we find broad disagreement between the participants. This diversity in judgments is needed to support the actual analysis. Indeed, a good theory of compositional vague

[4] For the adjective *complete*, we see that, despite the instructions, many participants gave a rating equal to the given degree of completion, instead of judging the applicability of the predicate. This can presumably be explained by our use of percentages in the stimuli, which map very easily onto the slider, since only this adjective showed such a pattern.

semantics should be able to explain as much of this individual variance on quantified cases as possible from the individual judgments on atomic cases. Having judgments on both atomic and quantified sentences spanning a large range from clearly true to clearly false is therefore crucial to properly evaluate such proposals.

3.5. Models Evaluation

We focus our attention on three flavors of fuzzy logics: Gödel, product, and Łukasiewicz (we will come back to other families of theories in the discussion). These logics differ by the t-norm they use to represent (strong) conjunction. Gödel logic uses the Zadeh operator min, product logic — the standard product among real numbers, and Łukasiewicz logic – the operator max$(0, x + y - 1)$. Fuzzy logics can be extended to first order, and this is usually done by assigning extending the weak conjunction/disjunction for ∀/∃, which always use the min / max operators (Hájek 1998). Note, however, that all t-norms are commutative and associative, and so we can easily build quantifiers from the different t-norms and test which match best the participants' behavior.

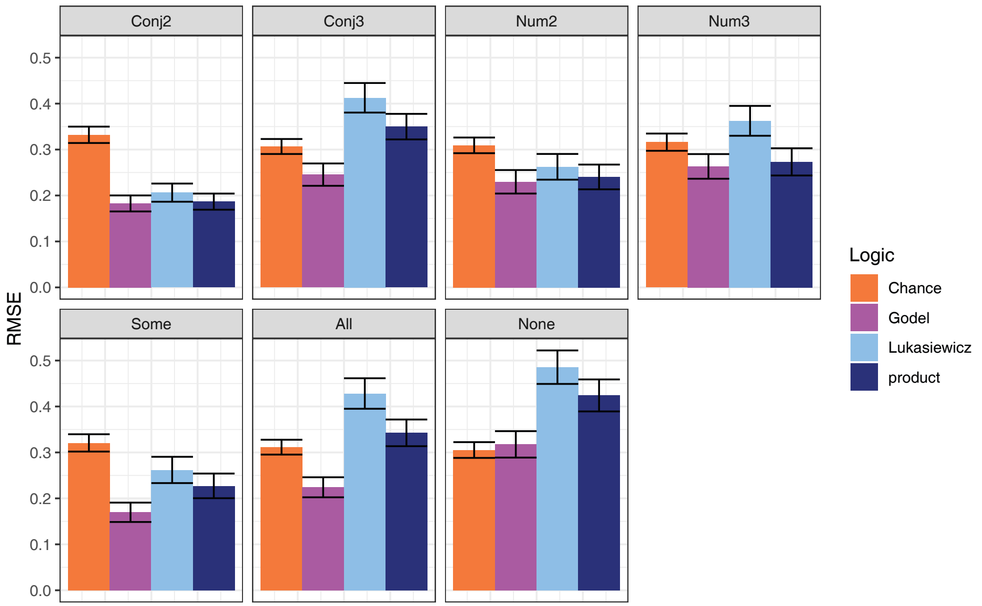

Figure 5. Root-mean-square error for each of the three fuzzy logics and the chance baseline, across quantifiers

One important caveat is that formulas which may be equivalent in classical logic might not be equivalent in different fuzzy logics. We therefore had to make some arbitrary choices on how to translate our quantifiers, and different choices would possibly yield different results. The main point of contention would be numerals, which we translated as disjunctions of conjunctions. This corresponds to an 'at least' interpretation, in line with most of the recent literature on the semantics of numerals (Spector 2013; Cremers et al. 2022, a.o.), but these 'at-least' interpretations are often strengthened into 'exactly' readings through implicatures, an aspect we leave aside here for simplicity. More detail on our translations and the resulting predictions is provided in Table 3.

Predictions for each logic are derived for each participant individually, from the ratings they gave to each item in the set. For instance if a participant judged that student A is 'tall' to a degree of 0.8 and student B to a degree of 0.6, Gödel logic would predict that this participant will assign the sentence "Students A and B are tall" a truth- degree of $\min(0.8, 0.6) = 0.6$. Meanwhile, the product logic would predict $0.8 \times 0.6 = 0.48$, whereas Łukasiewicz logic, $\max(0, 0.8 + 0.6 - 1) = 0.4$. In addition to the three logics described above, we also included a chance baseline, which assigns a value of 0.5 to every sentence. We then compare these predictions to the judgments actually expressed on these complex sentences, and compute the root-mean-squared error across all cases labelled 'borderline' (it was decided a priori to focus on these cases only as the predictions on extreme cases are very similar, and beating the chance level would be too easy on clear true and false cases). Figure 5 shows the results broken down by quantifier. From this, it is clear that, among the three fuzzy logics we tested, Gödel logic is the best predictor of the participants' behavior. Product and Łukasiewicz logics even perform worse than chance on some quantifiers, including on triple conjunctions, for which, there is no translation debate.

4. Discussion

The first and most obvious conclusion from our results is that, if one was to adopt a graded-truth approach to vagueness, they should use Gödel fuzzy logic over the alternatives we tested. It so happens that most work using fuzzy logic as a model for vagueness in the linguistic and philosophy literature seems to have made this choice, but it is not entirely clear whether it was intentional or simply following Zadeh (1975). Previous empirical work has already shown that Zadeh's operators offer a good empirical coverage of simple sentences (Hersh & Caramazza 1976, or, more recently, Leffel et al. 2019), but these studies did not test other fuzzy logics, and our results on simple conjunctions show that all the three tested logics perform equally well there, suggesting that the results of these previous studies would not have been sufficient to decide between them anyway. To the best of our knowledge, our experiment is therefore the first to justify the choice of Zadeh operators for natural language vagueness on the basis of quantitative data. Interestingly, there are other arguments that could favor other logics. Dubois and Prade (1980) prove that, for a fuzzy logic, excluded-middle laws are inconsistent with distributivity and idempotency of union and intersection. Unlike Gödel logic, Łukasiewicz logic follows the excluded-middle laws, albeit at the price of violating distributivity and idempotency, which Gödel operators do respect. Meanwhile, product logic satisfies neither excluded-middle nor idempotency. The choice of one logic over another is thus a deeper question than just finding the mathematical expressions that best describe participants' behavior, as it determines the key properties we assign to vague sentences.

Łukasiewicz logic also has some useful algebraic properties. McNaughton (1951) proved an important theorem stating that formulas of this logic denote an important class of functions, and this theorem has garnered attention recently because of its applications in AI. Indeed, the relevant class of functions corresponds to a class of artificial neural nets, and the theorem tells us that any such net can be represented by a formula in Łukasiewicz logic (see, e.g., Castro & Trillas 1998; Amato et al. 2002; Nola & Vitale 2020, for the initial results and extensions). This could get us closer to the Holy Grail of 'explainable AI', where, after training a neural net on some data, we can obtain a useful symbolic representation of what has been learned. Yet, our results suggest that the formulas obtained this way would have limited interpretability, as the connectives would not correspond to their natural human interpretation. The lack of distributivity and idempotency would also make it extremely difficult to simplify these formulae, notwithstanding the fact that no concrete implementation of these neural nets seems to have been proposed, thereby casting further doubt on their real-world usefulness.

Our discussion so far presupposes that we adopt a graded-truth approach to vagueness, but this approach has long fallen out of favor in the literature on vagueness, and for a few good reasons. The main arguments include an improper account of contradictions and tautologies (as we saw that Gödel logic fails to satisfy the laws of excluded- middle), and a failure to represent higher-order vagueness (see Williamson 1994; Osherson & Smith 1997 for more). Now, in the introduction, we have discussed another approach which could lead to quantitative predictions, namely, a probabilistic version of epistemicism, which has been particularly popular in the wake of Bayesian game-theoretic approaches to pragmatics (Lassiter & Goodman 2014, 2017; Qing & Franke 2014). Under a few reasonable assumptions, it can be shown that the predictions of this approach for the sentences tested in our experiment exactly match those of Gödel logic. Indeed, if the denotation of a vague predicate can be reduced to comparison with a fixed unknown threshold θ, the probability that two elements *a* and *b* of known measurement both exceed θ is simply the minimum of the probabilities for each element individually[5]:

$$\begin{aligned} P(\mu(a) > \theta \wedge \mu(b) > \theta) &= P(\max(\mu(a),\mu(b)) > \theta) \\ &= \left| \begin{array}{ll} P(\mu(a) > \theta) & \text{if } \mu(a) \geq \mu(b) \\ P(\mu(b) > \theta) & \text{if } \mu(b) \geq \mu(a) \end{array} \right. \\ &= \min(P(\mu(a) > \theta), P(\mu(b) > \theta)) \end{aligned}$$

To apply this result to our experiment, we need to assume three things: (i) our predicates are unidimensional (see D'Ambrosio & Hedden 2023 for a recent discussion of multidimensional adjectives), (ii) all items in the scope of the quantifier are compared to the same threshold (i.e., they all belong to the same comparison class), and (iii) the only source of uncertainty is the adjective itself. Assumption (iii) is trivial here since the measurements of every item on the relevant scale was given to participants, as well as the satisfaction

[5] This generalizes immediately to quantifiers 'every', 'some', and 'no', and we let the reader verify the case of numerals.

of the restrictor predicate[6]. Assumption (i) could be debated, but, except for ‘safe’, none of the adjectives we used seems multidimensional. Besides, the design of the experiment, where the participants were given measurements on a single dimension, was a strong cue towards a unidimensional interpretation. Regarding assumption (ii), we again find no reason to doubt it, as we avoided introducing any distinction within the relevant item. The only distinction we introduced in the course of the experiment was the predicate used for the restrictor, but we deem it unlikely that the participants would have used it to restrict the comparison class, as it would have resulted in trivial interpretations (e.g., “All black cars are powerful [for black cars]”).

This equivalence (on our results) has two consequences. The first is practical: our results are as compatible with a probabilistic interpretation of epistemicism as they are with a Gödel logic variant of graded-truth. The reader may ask why this would matter; after all, did the literature not already settle this debate 30 years ago? While this is true for a plain fuzzy logic approach to vagueness, several alternatives have tried to address its limitations while retaining some of its interesting properties (MacFarlane 2010 for ‘fuzzy epistemicism’, Smith 2008, 2011 for a ‘plurivaluationalist fuzzy’ approach). Similarly, our presentation of ‘probabilistic epistemicism’ above glosses over the fact that the probabilities themselves would not be precisely accessible to introspection. Once we have added this layer of higher-level uncertainty, distinguishing between the two approaches becomes trickier – but also more relevant.

The second consequence is conceptual: if, with other empirical results, we were able to decide in favor of the graded-truth approach, this equivalence would provide an answer as to *why* it is this fuzzy logic that best describes human behavior. Indeed, if, on simple cases, the Zadeh operators are equivalent to an (arguably more rational) probabilistic approach, these operators would provide a good heuristic to approximate probability calculations at a much lower computation cost (probabilities require keeping track of dependencies between propositions, whereas fuzzy logic only needs the assigned truth values). Human participants would therefore rely on Gödel logic as a shortcut to approximate the intended semantics on complex sentences.

So how could we tease apart graded-truth from epistemicism on empirical grounds? The first obvious answer would be to look at cases where the assumptions for the equivalence discussed above are not met: we could study multidimensional adjectives or, more conveniently, conjunctions of different adjectives (e.g., “Ann is tall and rich”). The thresholds for two unrelated adjectives would arguably be independent, so the predictions of probabilistic epistemicism would now match those of product logic rather than Gödel, while the graded-truth approach is blind to such dependencies, or lack thereof. Fortunately, while we leave this investigation for future research, there are already relevant results in the existing literature.

[6] Note that there could be uncertainty on the exact mapping from concrete measurements to degrees, but as long as this mapping is monotone increasing, this does not matter here. For now, we also disregard uncertainty on the probabilistic distribution of θ (i.e., higher-order vagueness), as it is orthogonal to the point discussed here.

First, Bonini et al. (1999) carried a number of experiments to distinguish between various trivalent theories, fuzzy logic, and epistemicism. While their main results only rule out some trivalent theories (namely, those that assign borderline cases a 'both-true-and-false' interpretation), they present further results demonstrating that the participants' behavior in cases of standard ignorance is similar to their behavior with vague predicates. While this is not definitive evidence for epistemicism, it at least demonstrates that observing gaps between the positive and negative extension of a predicate is not evidence for a non-binary denotation. Serchuk et al. (2011) criticize some aspects of the methodology, and fail to replicate the results with a slightly different design however. Another argument presented in Bonini et al. (1999) and elsewhere against fuzzy logic specifically comes from the law of excluded-middle and relies on sentences such as (3) and (4). In (3), Gödel logic would assign a truth degree close to 0.5 since the truth of each conjunct is close to 0.5, but the sentence feels intuitively false. Similarly, (4-a) and (4-b) would be assigned the same truth degree, but since we know that Tim is shorter than Jim, (4-b) should be clearly false. Crucially, this argument would still hold against the 'lifted' fuzzy approaches of Smith and MacFarlane discussed above.

(3) *Context: Two lamps emit light at the exact same wavelength, which is a borderline case of 'red'*

The lamp on the left is red, and the lamp on the right is not red

(4) *Context: Jim is borderline tall, such that he is as 'tall' as he is 'not tall', and Tim is slightly shorter than Jim.*
  a. Tim is tall and Jim is tall
  b. Tim is tall and Jim is not tall

While fuzzy logic fails to categorize sentences such as (3) and (4-b) as clearly false, epistemicism (and many other approaches) have been criticized for rejecting so-called 'borderline contradictions' too hastily. These are sentences such as (5). Ripley (2011) and Alxatib and Pelletier (2011) find that these sentences receive a moderate acceptability when Mary is borderline tall. While the authors of these studies favor trivalent approaches, the results align even better with fuzzy logic accounts, and would definitely be challenging for an epistemic account.

(5) a. Mary is neither tall nor not tall
  b. Mary is tall and not tall

Finally, Cremers et al. (2017) tested how naïve participants assign probabilities to various sentences which have been claimed to display truth-value gaps. For vague sentences, they found that the majority of the participants behave classically (assigning to 'x is big' and 'x is not big' probabilities which add up to 1), but a 25% minority still treated the borderline cases as truth gaps (assigning probabilities which summed to less than 1). How to link this observation to accounts of vagueness which are primarily focused on truth and assertability is not immediately clear, however.

5. Conclusions

To sum up, there are rather strong arguments on each side of the debate, but both sides have found solutions to accommodate the problematic data, and so this debate has not been definitively settled yet. How, then, could we reconcile the apparent falsity of (3) on the one hand, and the relative tolerance towards (5)? Graded-truth advocates such as MacFarlane argue that examples like (3) are degraded for pragmatic reasons (e.g., they are certain to never receive a high degree of truth, and this makes them unassertable). Conversely, epistemicists would argue that (5) is relatively acceptable for pragmatic reasons (e.g., involving meta-linguistic negation). Cobreros et al. (2012) propose the Tolerant/Classical/Strict semantics in part to account for such cases, with the idea that different occurrences of the same predicate may receive different interpretations.

At this point, it would probably be good to take a step back and think about what exactly we are trying to model when we model 'vagueness'. The original literature was primarily focused on semantics, but, as the field progresses and subtler empirical arguments are put forward, we see that we are more and more dealing with questions of assertability, and, therefore, pragmatics (see also Canonica 2022, for a discussion of the need to disentangle semantic and pragmatic factors in experimental settings). In this context, probabilistic game-theoretic models may become particularly valuable because they allow us to make explicit our assumptions about semantics, pragmatics, and their interactions. As explained in the introduction, most accounts see vagueness as primarily pragmatic, with semantics only providing free variables (Lassiter & Goodman 2017), but some assume a split between semantics and pragmatics (such as the pragmatic prototype theory in van Tiel et al. 2021). Crucially, such models could allow us to properly assess the claims from one or the other side that problematic examples can be explained by pragmatics. Cremers (2022b) proposes a general model of pragmatic reasoning on vague sentences, which aims to explain interactions between vagueness and implicatures. This account remains agnostic with respect to the treatment of first-order vagueness, which can either be relegated to semantics, adopting a graded-truth approach, or kept as part of pragmatics, remaining closer to the standard epistemic approach. Second-order vagueness, however, must be probabilitistic and pragmatic in this model. Crucially, the assertability is not simply proportional to the probability that the sentence is true, but, following Spector (2017), to the expected log-probability that a listener would retrieve the correct interpretation after hearing the sentence. This non-linearity is crucial in explaining why borderline contradictions are not usually assertable, and why some implicatures disappear with vague terms (which is the focus of the paper). A promising avenue for future research, then, would be to revisit cases like (3) or (5), and test which semantics is needed to best account for them.

References

Alxatib, S., & Pelletier, J., 2011. On the Psychology of Truth-Gaps. In *Vagueness in Communication*, 13–36). Springer.

Amato, P., Di Nola, A., & Gerla, B., 2002. Neural Networks and Rational Łukasiewicz logic. In *2002 Annual Meeting of the North American Fuzzy Information Processing Society Proceedings.* NAFIPS-FLINT *2002 (cat. no. 02th8622)*, 506–510. doi: 10.1109/NAFIPS.2002.1018111

Bonini, N., Osherson, D., Viale, R., & Williamson, T., 1999. On the Psychology of Vague Predicates. *Mind & Language*, 14(4): 377–393.

Burnett, H., 2014. A Delineation Solution to the Puzzles of Absolute Adjectives. *Linguistics and Philosophy*, *37*(1): 1–39. doi: 10.1007/s10988-014-9145-9

Canonica, C. R., 2022. *Borderline Contradictions* (unpublished doctoral dissertation). UCL (University College London).

Carcassi, F., Steinert-Threlkeld, S., & Szymanik, J., 2021. Monotone Quantifiers Emerge via Iterated Learning. *Cognitive Science*, *45*(8), e13027. doi: https://doi.org/10.1111/cogs.13027

Castro, J. L., & Trillas, E., 1998. The Logic of Neural Networks. *Mathware and Soft Computing*, *5*: 23–37.

Cobreros, P., Egré, P., Ripley, D., & van Rooij, R., 2012. Tolerant, Classical, Strict. *Journal of Philosophical Logic*, 41(2): 347–385. doi: 10.1007/s10992-010-9165-z

Cremers, A. 2022a. Interpreting Gradable Adjectives: Rational Reasoning or Simple Heuristics? In G. Bîlbîie, B. Crysmann, & G. Schaden (Eds.), *Empirical Issues in Syntax and Semantics* 14: 31–60. Paris: CSSP.

Cremers, A., 2022b. A Rational Speech-Act Model for the Pragmatic Use of Vague Terms in Natural Language. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of CogSci 44*, 149–155.

Cremers, A., 2024. Vagueness and Pragmatic Reasoning in Quantified Sentences. In *Proceedings of the 24th Amsterdam Colloquium.*

Cremers, A., Coppock, L., Dotlačl, J., & Roelofsen, F., 2022. Ignorance Implicatures of Modified Numerals. *Lin- guistics and Philosophy*, 45(3): 683–740. doi: 10.1007/s10988-021-09336-9

Cremers, A., Križ, M., & Chemla, E., 2017. Probability Judgments of Gappy Sentences. In S. Pistoia-Reda & F. Domaneschi (eds.), *Linguistic and PsycholinguisticAapproaches on Implicatures and Presuppositions,* 111–150. Springer.

D'Ambrosio, J., & Hedden, B., 2023. Multidimensional Adjectives. *Australasian Journal of Philosophy*, 102(2):_1–25.

Denić, M., & Szymanik, J., 2022. Are *Most* and *More than Half* Truth-Conditionally Equivalent? *Journal of Semantics*, 39(2): 261–294.

Douven, I., 2016. Vagueness, Graded Membership, and Conceptual Spaces. *Cognition*, 151: 80–95. doi: https://doi.org/10.1016/j.cognition.2016.03.007

Douven, I., 2021. Fuzzy Concept Combination: An Empirical Study. *Fuzzy Sets and Systems*, 407: 27–49.

Dubois, D., & Prade, H., 1980. New Results about Properties and Semantics of Fuzzy Set-Theoretic Operators. In P. P. Wang & S. K. Chang (eds.), *Fuzzy Sets: Theory of Applications to Policy Analysis and Information Systems*p, 59–75. Springer.

Fine, K., 1975. Vagueness, Truth and Logic. *Synthese*, 30(3–4): 265–300.

Glöckner, I., 2008. *Fuzzy Quantifiers: a Computational Theory,* Vol. 193. Springer.

Hájek, P., 1998. *Metamathematics of Fuzzy Logic*. Springer Dordrecht.

Halldén, S., 1949. The Logic of Nonsense. In *Uppsala Universitets a˚rsskrift* (Vol. 15, p. 132). Wiesbaden: Harrassowitz Verlag.

Hersh, H. M., & Caramazza, A., 1976. A Fuzzy Set Approach to Modifiers and Vagueness in Natural Language. *Journal of Experimental Psychology: General*, 105(3): 254-276.

Kennedy, C., 1999. *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. New York: Routledge. doi: 10.4324/9780203055458

Kennedy, C., & McNally, L., 2005. Scale Structure, Degree Modification, and the Semantics of Gradable Predicates. *Language*, 81(2): 345–381.

Kotek, H., Sudo, Y., & Hackl, M., 2015. Experimental Investigations of Ambiguity: the Case of Most. *Natural Language Semantics*, 23(2): 119–156.

Lakoff, G., 1973. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic*, 2(4): 458–508.

Lassiter, D., & Goodman, N. D., 2014. Context, Scale Structure, and Statistics in the Interpretation of Positive-Form Adjectives. In T. Snider (Ed.). In *Proceedings of SALT* 23, 587–610). Washington, D.C.: LSA.

Lassiter, D., & Goodman, N. D., 2017. Adjectival Vagueness in a Bayesian Model of Interpretation. *Synthese*, 194(10): 3801–3836. doi: 10.1007/s11229-015-0786-1

Leffel, T., Cremers, A., Gotzner, N., & Romoli, J., 2019. Vagueness in Implicature: The Case of Modified Adjectives. *Journal of Semantics*, 36(2): 317–348. doi: 10.1093/jos/ffy020

MacFarlane, J., 2010. Fuzzy Epistemicism. In R. Dietz & S. Moruzzi (eds.), *Cuts and Clouds: Vagueness, its Nature and its Logic,* 438–463. Oxford: Oxford University Press.

Marty, P., Chemla, E., & Sprouse, J., 2020. The Effect of Three Basic Task Features on the Sensitivity of Acceptability Judgment Tasks. *Glossa: a Journal of General Linguistics*, 5(1): 72.

McNally, L., 2011. The Relative Role of Property Type and Scale Structure in Explaining the Behavior of Gradable Adjectives. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), *Vagueness in Communication*, 151–168). Berlin, Heidelberg: Springer Berlin Heidelberg.

McNaughton, R., 1951. A Theorem about Infinite-Valued Sentential Logic. *The Journal of Symbolic Logic*, 16(1): 1–13.

Nola, A. D., & Vitale, G., 2020. Łukasiewicz Logic and Artificial Neural Networks. In O. Kosheleva, S. P. Shary, G. Xiang, & R. Zapatrin (eds.), *Beyond Traditional Probabilistic Data Processing Techniques: Interval, Fuzzy etc. Methods and their Applications*, 137–149. Cham: Springer International Publishing. doi: 10.1007/978-3-030- 31041-7 8

Osherson, D., & Smith, E. E., 1997. On typicality and vagueness. *Cognition*, 64(2): 189–206.

Pietroski, P., Lidz, J., Hunter, T., & Halberda, J., 2009. The Meaning of 'Most': Semantics, Numerosity and Psychology. *Mind & Language*, 24(5): 554–585.

Qing, C., & Franke, M., 2014. Gradable Adjectives, Vagueness, and Optimal Language Use: A Speaker-Oriented Model. In T. Snider, S. D'Antonio, & M. Weigand (eds.), *Proceedings of SALT 24,* Vol. 24, 23–41. Washington, D.C.: LSA.

Ramotowska, S., Haaf, J., Van Maanen, L., & Szymanik, J., 2024. Most Quantifiers Have Many Meanings. *Psycho-Nomic Bulletin & Review*. doi: 10.3758/s13423-024-02502-7

Ripley, D., 2011. Contradictions at the Borders. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (eds.),

Vagueness in communication, 169–188. Springer.

Rosch, E. H., 1973. On the Internal Structure of Perceptual and Semantic Categories. In T. E. Moore (ed.), *Cognitive Development and Acquisition of Language,* 111–144). San Diego: Academic Press. doi: 10.1016/B978-0-12- 505850-6.50010-4

Russell, B., 1923. Vagueness. *The Australasian Journal of Psychology and Philosophy*, 1(2): 84–92.

Sarafoglou, A., Giacobello, A., Godmann, H., Johnson, T., Visser, I., Haaf, J. M., & Szymanik, J. (2024, Oct). *A bayesian framework to study individual differences in semantic representations.* OSF Preprints. doi: 10.31219/osf.io/hvgb7

Schöller, A., & Franke, M., 2017. Semantic Values as Latent Parameters: Testing a Fixed Threshold Hypothesis for Cardinal Readings of Few & Many. *Linguistics Vanguard*, 3(1): 2016–0072. doi:10.1515/lingvan-2016-0072

Serchuk, P., Hargreaves, I., & Zach, R., 2011. Vagueness, Logic and Use: Four Experimental Studies on Vagueness. *Mind & Language*, 26(5): 540–573.

Smith, N. J., 2008. *Vagueness and Degrees of Truth*. Oxford University Press.

Smith, N. J., 2011. Fuzzy Logic and Higher-Order Vagueness. *Understanding Vagueness: Logical, Philosophical and Linguistic Perspectives*, 1–19.

Solt, S., 2016. On Measurement and Quantification: The Case of *Most* and *More than Half*. *Language*, 92(1): 65–100.

Spector, B., 2013. Bare Numerals and Scalar Implicatures. *Language and Linguistics Compass*, 7(5): 273–294.

Spector, B., 2017. The Pragmatics of Plural Predication: Homogeneity and Non-Maximality within the Rational Speech Act Model. In A. Cremers, T. van Gessel, & F. Roelofsen (eds.), *Proceedings of the 21st Amsterdam Colloquium,* 435).

Sprouse, J., Schütze, C. T., & Almeida, D., 2013. A Comparison of Informal and Formal Acceptability Judgments Using a Random Sample from Linguistic Inquiry 2001–2010. *Lingua*, 134: 219–248.

Tye, M., 1994. Sorites Paradoxes and the Semantics of Vagueness. *Philosophical Perspectives*, 8: 189–206.

van Tiel, B., Franke, M., & Sauerland, U., 2021. Probabilistic Pragmatics Explains Gradience and Focality in Natural Language Quantification. *Proceedings of the National Academy of Sciences*, 118(9), e2005453118. doi: 10.1073/pnas.2005453118

Williamson, T., 1994. *Vagueness*. London: Routledge. doi: 10.4324/9780203014264

Xiang, M., Kennedy, C., Xu, W., & Leffel, T., 2022. Pragmatic Reasoning and Semantic Convention: A Case Study on Gradable Adjectives. *Semantics and Pragmatics*, 15: 9–EA.

Zadeh, L. A., 1975. Fuzzy Logic and Approximate Reasoning. *Synthese*, 30(3): 407–428.

Appendix: Tables

Table 1. Vague adjectives and restricting predicates used in the experiment. The first 6 adjectives (tall—large) are relative adjectives, the next three (late–profitable) are minimum-standard absolute adjectives, while the last three (safe– full) are maximum-standard absolute adjectives

adjective	items	dimension	predicate
tall	students	height	drink coffee have roommates play video games use public transport
powerful	cars	horsepower	have heated seats are black have a touchscreen have leather interiors
young	employees	age	use black pens are left-handed keep journals have houseplants
hot	days	temperature	were workdays had high humidity were windy had heavy traffic
expensive	cars	price	have reflector headlights have tinted windows have the gas tank door on the left side are grey
large	apartments	surface	are located on the top floor come pre-furnished have double-glazed windows have air conditioning
late	employees	time	drink tea work in HR use post-it notes bring packed lunches
spicy	dishes	spiciness	contain lemongrass come with complimentary prawn crackers contain holy basil are served with noodles
profitable	companies	profit	have employee uniforms have 24/7 customer support are construction companies are family businesses
safe	neighborhoods	violent crime rate	have a public library have a hardware store are situated North of the city center border a river

adjective	items	dimension	predicate
complete	investigations	completion	are about credit card fraud require an official statement from the bank involve a repeat offender affect retired clients
full	flights	empty seats	included a layover left in the morning used a Boeing model plane were international

Table 2. Design of the four cases for each quantifier: we describe each case by the number of true/borderline/false items included in the restrictor. The true and false cases were taken from each end of the scale. The borderline cases were identified as the two items with the rating closest to 50% in a previous experiment which used the same stimuli

Quantifier	Case	#False	#Borderline	#True
Conj2	False	1	0	1
	T1	0	1	1
	T2	1	1	0
	True	0	0	2
Conj3	False	2	0	1
	T1	0	1	2
	T2	1	2	0
	True	0	0	3
Two	False	2	1	0
	T1	2	1	2
	T2	2	1	1
	True	2	0	2
Three	False	2	2	0
	T1	1	1	3
	T2	1	2	1
	True	1	0	3
Some	False	2	0	0
	T1	0	2	2
	T2	2	2	0
	True	1	1	2
All	False	2	1	1
	T1	0	2	2
	T2	0	2	1
	True	0	0	3
None	False	2	0	1
	T1	2	1	0
	T2	2	2	0
	True	3	0	0

Table 3. Translations and resulting predicted truth-value from each logic

(a) Translation for each target sentence

Quantifier	**Formula**
Conj2	$A \otimes B$
Conj3	$A \otimes B \otimes C$
Some	$\bigoplus X$
All	$\bigotimes X$
None	$\neg \bigoplus X$
Two	$\bigoplus_{x \neq y \in X} x \otimes y$
Three	$\bigoplus_{x,y,z \text{ distinct}} x \otimes y \otimes z$

(b) Corresponding truth-values in each different logic

	Gödel	**product**	**Łukasiewicz**
Conj2	$\min(A, B)$	$A \times B$	$\max(0, A + B - 1)$
Conj3	$\min(A, B, C)$	$A \times B \times C$	$\max(0, A + B + C - 2)$
Some	$\max X$	$1 - \Pi(1 - \mathrm{X})$	$\min(1, \Sigma X)$
All	$\min X$	ΠX	$\max(0, \Sigma X - \#X + 1)$
None	$1 - \max X$	$\Pi(1 - X)$	$\max(0, 1 - \Sigma X)$
Two	$\max_{(x,y)} \min(x, y)$	$1 - \Pi(1 - x \times y)$	$\min(1, \Sigma(x + y - 1))$
Three	$\mathrm{Max}_{(x,y,z)} \min(x, y, z)$	$1 - \Pi(1 - x \times y \times z)$	$\min(1, \Sigma(x + y + z - 2))$