Title
When is Likely Unlikely: Investigating the Variability of Vagueness

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Persaud, Kimele
McMahan, Brian
Alikhani, Malihe
et al.

Publication Date
2017

Peer reviewed

When is Likely Unlikely: Investigating the Variability of Vagueness

Kimele Persaud (kimele.persaud@rutgers.edu)[*]

Brian McMahan (brian.mcmahan@rutgers.edu)[**]

Malihe Alikhani (ma1195@cs.rutgers.edu) [**]

Kevin Pei (kevin.pei@rutgers.edu) [*]

Pernille Hemmer (pernille.hemmer@rutgers.edu)[*]

Matthew Stone (matthew.stone@rutgers.edu)[**]
Department of Psychology, 152 Frelinghuysen Road, Piscataway, NJ 08854 USA[*]
Department of Computer Science, 96 Frelinghuysen Road, Piscataway, NJ 08854 USA[**]

Abstract

An important part of explaining how people communicate is to understand how people relate language to entities in the world. In describing measurements, people prefer to use qualitative words like 'tall' without precise applicability conditions, also known as *vague* words. The use of vague language varies widely across contexts, individuals, and tasks (single reference vs. comparisons between targets), but despite this variability, is used quite successfully. A potential strategy for using vague language is to leverage the set of alternative descriptors to settle on the best option. To determine whether people use this strategy, we conducted an experiment where participants picked vague words from sets of alternatives to describe either probability or color values. We varied the set of alternatives from which participants could choose. Empirical evidence supports the hypothesis that people use the set of available options to pick vague descriptors. The theoretical implications of this work are discussed.

Keywords: Vagueness, Alternative sets, Probability, Color

Introduction

An important aspect of human communication is understanding how people use language to describe the world (Quine, 1969). In the simplest case, people use language to refer to concrete real world objects and categories, such as trees or humans. More interestingly, people use words such as *blue*, *tall*, and *likely* to flexibly refer to indefinite ranges of continuous values across different contexts—a phenomenon known as *vagueness*. Vague words vary in evoking degrees along different kinds of dimensions (Kennedy, 2007; Kennedy & McNally, 2010). This paper explores people's expectations for different classes of vague words and how they leverage these expectations to communicate effectively.

The constrained variability of vague language

Although the use of vague language is ubiquitous in everyday talk, it varies dramatically across communicative situations (Budescu & Wallsten, 1987). To start, different people relate vague words differently to the values they want to describe (Budescu & Wallsten, 1987; Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986). For example, in comparable contexts, different people may use the same word to describe different values, or different words to describe the same values. Nevertheless, meanings in context are not random: vague words seem to always denote bounded, convex regions in the appropriate property space (Gärdenfors, 2004). For example, we can appeal to the convexity of color categories to explain patterns of color naming within and across language communities (Jäger, 2010; Regier, Kay, & Khetarpal, 2007).

At the same time, individuals' use of vague language varies as a function of the context. People use vague words differently depending on the specific objects they need to distinguish in a situation (Van Deemter, 2006). They also use vague words differently depending on the possible alternative descriptions that would be appropriate (Degen, 2015). For instance, the presence of numbers in the set of available options (e.g.['some', 'all', 'not all', '4']), influences the use of the option 'some' (Degen & Tanenhaus, 2015). Again, there are limits on such effects. For example, absolute terms, such as 'empty', 'flat' and 'straight', are more constrained in how they vary in context than terms that signal open-ended comparisons, such as 'tall' (e.g. Leffel, Xiang, & Kennedy, 2016).

Finally, vague language varies as a function of speakers' semantic memory, as revealed by the implicit class to which comparisons are drawn (Lassiter, 2009; Wallsten et al., 1986; Kennedy, 2007). In the quintessential example, the word *tall* is understood very differently when used to describe a basketball player versus a toddler, and differently again when used to describe a skyscraper versus a glass of water (Schmidt, Goodman, Barner, & Tenenbaum, 2009). Similar semantic effects are seen in the relationships between different cultures, environments, and their color categories (Regier, Carstensen, & Kemp, 2016; Stickles, 2014).

Despite this constrained variability, people generally communicate successfully with vague words and prefer to use vague language in many tasks (Van Deemter, 2012).

Characterizing the variability of vagueness

Bayesian cognitive modeling suggests accounting for these effects in terms of expectations derived from semantic memory, the communicative context and patterns of individual variation (Potts, Lassiter, Levy, & Frank, 2016). Conversely, it suggests we can also characterize people's semantic and pragmatic representations through analysis of their interpretation of vague language.

In particular, it's natural to suppose that language users coordinate on specific interpretations in context by assuming that speakers have chosen the most informative description from the available alternatives in light of their expectations. Although several models predict that interpretations of vague language will vary in this way (Potts et al., 2016; Wallsten et al., 1986), empirical evidence supporting this hypothesis is thin.

To investigate this hypothesis, we conducted a forced choice experiment where subjects selected one of a small number of linguistic alternatives to describe a value. Critically, as we varied the number of alternative descriptions, we provided a reject option (i.e. 'none of the above') to measure limits in flexibility of vague terms. Despite a large range of related work, no previous studies have addressed this issue explicitly. We compared performance in the forced choice experiment to a free generate task where participants provided their own words to describe values.

We hypothesize that two possible behaviors may arise. When presented with different sets of alternative descriptions, people may flexibly use the same vague word to refer to different ranges of values. In contrast, people may consistently use vague words to describe the same values regardless of the alternative set of words, instead choosing rejection when preferred options are unavailable.

To foreshadow, we find evidence for both patterns, depending on semantic domain. Probability descriptions vary widely depending on the alternative terms presented to subjects, but color terms vary much less. These results suggest that, although individuals representations of the meaning of some vague words are broadly stable, speakers do adjust boundaries within the available range to give the most information in context. [1] We outline empirical and theoretical consequences of this finding for future work, emphasizing the need to characterize individual differences and contextual variation jointly, as well as the need to explicitly contrast speaker models based on strategic and heuristic choice.

Experiment

To assess the role of the set of available terms in constraining vague language, we elicit labeling behavior in a task where participants are shown a property value (i.e. probability and hue), and are asked to either choose a corresponding label from a given set of options (N-AFC), or freely generate a label that corresponds to the presented value. In the N-AFC cases, we expect that the distributions will reflect peoples willingness (or lack of willingness) to "stretch" their category assignment of values based on available terms. In the free generate cases, we expect that the distributions will reflect people's natural tendencies of assigning terms to values.

[1] In fact, models of informativeness are often operationalized in terms of ruling out competing referential interpretations (Frank & Goodman, 2012). The only way to apply such models to our experiment is with the trivial assumption that all true descriptions are equally informative.

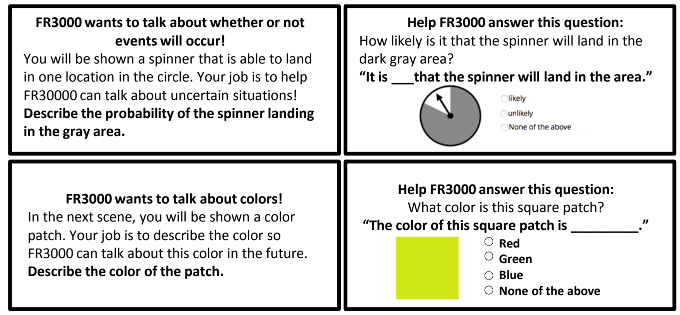

Figure 1: Sample stimuli for the two tasks: Probability (left two panels) and color (right two panels).

Methods

Participants Three-hundred and sixty individuals from the Amazon Mechanical Turk research pool participated in this study for monetary compensation.

Materials *Color.* The stimuli for the color condition consisted of 60 equally spaced values sampled from the winHSV240 (hue, saturation, and value) color space. The colors varied along the full range of the hue dimension, while saturation and value were held constant at 90%.[2] The set of available vague color words included seven of the eleven universal color terms (red, orange, yellow, green, blue, purple, and pink; Berlin & Kay, 1969). To create different conditions, we incrementally increased the number of color words available for participants to choose from - starting with three terms and ending with seven terms. We also included a free generate (see Table 1) condition resulting in six conditions for color in total. In addition to the AFCs for each condition, there was also a reject option, indicating that the color value was not described by any of the available color words. The color space was stratified into six regions, so that each participant only saw one stimuli from each region at equal intervals.This design ensured that participants were presented with values that spanned the entire property range.

Probability. The stimuli for the probability conditions also consisted of 60 equally spaced probability values on the range of 0-1. The vague probability words that could be used to describe the values were selected from a norming phase with a separate set of participants (N=32). The norming participants were simply asked to provide labels for randomly generated probability values. The six most frequently generated terms were then used here. To match the structure of the color task, the probability task was also comprised of six condi-

[2] Saturation and value were held constant to reduce the dimensionality of color space and to understand how people use color words to partition the range of hue values. This procedure is common practice when assessing expectations for basic color categories (Persaud & Hemmer, 2016; Sims, Ma, Allred, Lerch, & Flombaum, 2016)

Table 1: *List of Available Vague Words by Condition*

Condition		Probability		Color
1	[UL]	unlikely, likely	[3-TERM]	red, green, blue
2	[ULV]	unlikely, likely, very unlikely, very likely	[4-TERM]	red, green, blue, yellow
3	[ULS]	unlikely, likely, somewhat unlikely, somewhat likely	[5-TERM]	red, green, blue, yellow, purple
4	[VS]	very unlikely, very likely, somewhat likely, somewhat unlikely	[6-TERM]	red, green, blue, yellow, purple, orange
5	[ULVS]	unlikely, likely, very unlikely, very likely, somewhat unlikely, somewhat likely	[7-TERM]	red, green, blue, yellow, purple, orange, pink
6	[FG]	free generate	[FG]	free generate

tions ranging from two alternative forced choices (AFC) to six AFCs, and a free generate condition (see Table 1). The option to reject was present in all conditions. The probability space was stratified into six regions, so that each participant only saw one stimuli from each region at equal intervals.

Procedure Participants were told that they would be helping a robot to understand the meaning of vague words by assigning the words to different property values. Participants were first presented with a set of instructions describing the stimuli and the task. For the probability task, they were shown a pie chart with an arrow called a spinner and a shaded region denoting a probability value (Figure 1, top panels). They were informed that their job was to either pick from a given set of words or generate a word that described the likelihood of the spinner landing in the shaded region of the pie. For the color task, participants were presented with a single color patch and were asked to either choose one of the given color words or generate a word to describe the color Figure 1, bottom panels). Each set of instructions was provided immediately before the task that they described. Each participant described 12 unique property values (6 probability values and 6 hue values). The conditions and presentation order of values were randomized across participants.

Results

Probability Results We assessed whether or not participants consistently used the probability words to describe the same probability values across conditions (see Figures 2 and 3, left column), via linear mixed-effects models (lme) using the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). LME models test for significant differences in responses within experimental groups of primary interest (i.e. condition), while accounting for variability that results from factors that are experimentally uncontrolled (e.g. subjects). We followed up the modelling with planned pair-wise comparisons between conditions using a Tukey post hoc analysis, which corrects for family-wise error rates. In the LME models, subjects and stimuli order were always treated as random effects, while condition was treated as fixed. Probability and hue values were treated as the dependent measure and condition was treated as an indicator variable. For each probability word, we used a single lme model and compared it to the null. We started with a null model of participants and stimuli order and then added condition as a predictor in the alternate models. The null model predicts no difference in the assignment of probability words to values across conditions and the alternate models predict the opposite. Model fit was assessed using a likelihood-ratio test to compare the hypotheses of the null and alternate models. Condition was significant for probability words:*likely* ($\beta = 66.33$, $SE = 3.01$), *unlikely* ($\beta = 21.03$, $SE = 2.84$), and *somewhat likely* ($\beta = 47.67$, $SE = 4.72$). Model comparisons for each of these words favored the alternate models (*likely*: $\chi^2(4) = 36.77, p < .0001$, *unlikely*: $\chi^2(4) = 26.89$, $p < .0001$, and *somewhat likely*: $\chi^2(4) = 9.41$, $p = .02$). Planned pairwise comparisons were conducted to identify the conditions where the probability values differed for each word. For readability, we use codes to refer to the specific conditions (See Table 1 for the condition codes and probability terms available in each condition). Results showed a significant difference in the mean values for *likely* in the UL and ULV conditions ($p < .001$); ULV and ULS conditions ($p < 0.001$); and ULS and VS conditions ($p = 0.02$). See Figure 4, left panel, for a visualization of the cumulative changes in values for *likely* across AFC conditions.

Comparisons show that the probability values assigned to *unlikely* differed in the ULV and FG conditions ($p < 0.01$); UL and ULV conditions ($p = 0.01$); and ULS and ULV conditions ($p < 0.001$). See Figure 4, right panel, for a visualization of the cumulative changes in values for *unlikely* across AFC conditions. A difference in mean values for *somewhat likely* was observed in the ULVS and VS conditions ($p < 0.01$); and a marginal difference inULVS and FG conditions ($p = .055$). We also calculated the percentage of reject option responses in each N-AFC probability condition. In the order of Table 1, the reject option constituted 4%, 1%, 1%, 1%, and 1% of the responses. Taken together, the results suggest that not only are the assignment of probability words to values influenced by the set of alternative descriptions that could have been used, but also that the space of probability values do not have a strict partitioning. In other words, a varying number of probability terms can be flexibly used to

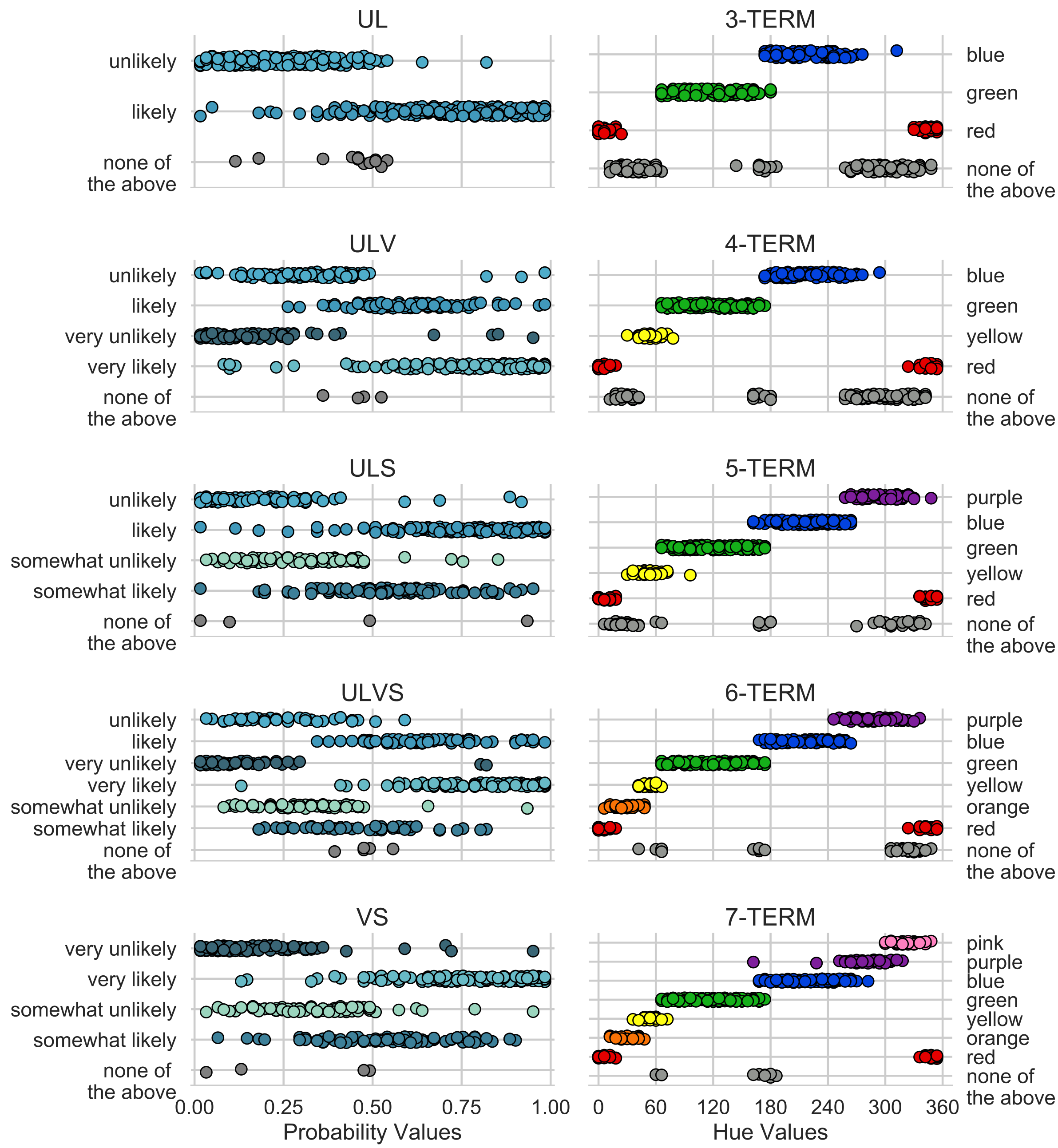

Figure 2: Responses from the N-AFC conditions. Labels on the y-axis were the options available to the participants and the x-axis shows the stimuli values. LEFT PANEL: For each plot, the probability terms selected vs. probability values presented to participants. RIGHT PANEL: For each plot, the color terms selected vs. the hue values presented to participants.

describe different values in probability space.

Color Results The most notable difference between the two domains is that the reject option was selected at a much higher rate for color than probability. In the order of Table 1, the reject option constituted 36%, 31%, 14%, 9%, and 3% of the responses for each N-AFC color condition. Like in the probability analysis, we used linear mixed-effects models to assess whether participants consistently used color words to describe hue values across conditions (see Figures 2 and 3, right column). Interestingly, model comparisons only favored the alternate model for the color *purple*. Again for readability, we use codes to refer to the specific conditions (see Table 1). Planned pairwise comparisons revealed the conditions where hue values for purple were different: 5-TERM and FG conditions ($p < .01$) and 5-TERM and 7-TERM conditions

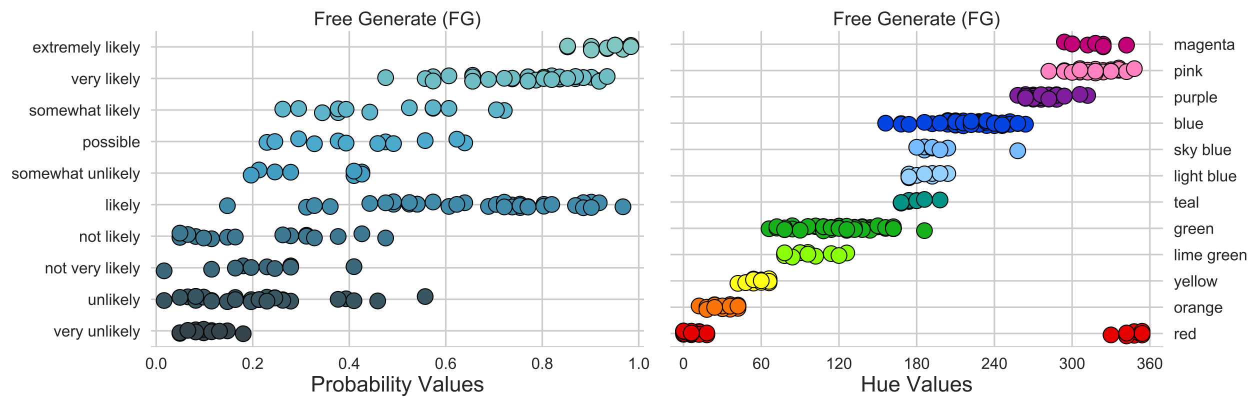

Figure 3: Responses from the free generate conditions. The y-axis shows the set of labels freely generated more than 5 times and the x-axis show the presented values. Probability responses are presented in the left panel and color responses are presented on the right.

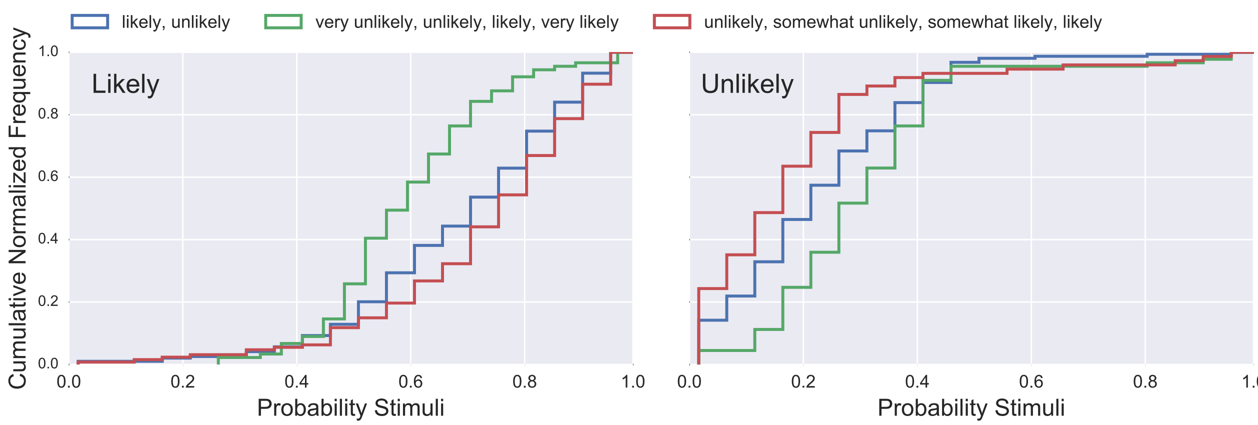

Figure 4: *Cumulative frequency curves for* likely *and* unlikely *across the AFC conditions. The curves show the relative rates of people using* likely *(on the left) and* unlikely *(on the right) probability terms given the alternative sets of probability words in each condition.*

($p < .01$). Importantly, there were no significant differences in the mean hue values for the remaining colors across conditions (i.e. red, green, blue, yellow, orange, and pink). The results suggest that, unlike some probability words, the assignment of color words to hue values are inflexible and are not influenced by the set of alternative descriptions that were offered to participants. Instead, participants assignments reflect that their preference for color categories already takes into account an alternative set of other color categories. This is further supported by the high rates of the reject option use in the conditions with fewer options for color descriptions.

Discussion

In this paper, we investigated how people assign vague words to probability and color values as a function of the set of available alternatives. We measured this behavior in two tasks where participants either selected a vague word from a fixed set or freely generated a word to describe values. Results revealed two interestingly opposing behaviors for probability and color.

For probability, words varied in their assignment to probability values when other vague terms were available. For example, *likely* was assigned to a different set of probability values in the UL condition where only *unlikely* was available, relative to the ULV condition where *very likely* and *very unlikely* were also available or the ULS condition where *somewhat likely* was available. In contrast, for color, the assignment of vague words to color values was much more rigid. In fact, *purple* was the only color that varied across conditions. Other words had relatively well-defined categories that did not overlap.

The results suggest that for probability, people are adopting the strategy of using the set of available terms to constrain variability. This is consistent with the well-known framing effect (Tversky & Kahneman, 1986) where decisions/preferences change as a function of how options are presented. For color, however, they are not adopting this strategy.

There are two potential reasons why this may be. First, the set of color words used in the task are already constrained to the basic universal color categories. It is possible that people are more flexible when the vague color word is not drawn from the 11 universal terms. For example, *teal*, which was a freely generated response might shift in its assignment to color values depending on the available options. If blue is present in the set, *teal* might be selected for more greenish hues, and if green is in the set, *teal* might be selected for bluish hues. Alternatively, it could be that probability words encode a relative comparison in a way that color words do not (Leffel et al., 2016), and this semantic difference stabilizes the interpretations of color words in context. In other words, color terms come with an intrinsic range of applicability, not just a prototypical or ideal instance of the term.

A possible limitation of this work is in the finding that the distributions for most probability terms are very broad and overlapping (See Figure 2), which might be due to either individual differences or the context provided by the alternative sets manipulation. The current methodology is insufficient to distinguish between these. One way we could assess this would be to build generative statistical models which simulate the behavior of the participants under the two possible stories and compare the simulations to the empirical data. At the same time, this work lays the groundwork for examining future questions such as: how do you represent the applicability of words like *likely* in ways that explain their constrained variability? And how do speakers combine their sense of what would be a good description with factors like their expectations about how a description will be interpreted?

Acknowledgements

This work is based upon a project supported by the National Science Foundation grant 1526723, the National Science Foundation Career grant number 1453276, and partially supported by National Science Foundation Graduate Research Fellowship under Grant Number NSF DGE 0937373.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their university and evolution*. California UP.

Budescu, D. V., & Wallsten, T. S. (1987). Subjective estimation of precise and vague uncertainties. *Judgmental forecasting*, 63–81.

Degen, J. (2015). Investigating the distribution of *some* (but not *all*) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, 1–55.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, 667–710.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 998–998.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*.

Jäger, G. (2010). Natural color categories are convex sets. In *Logic, language and meaning*. Springer.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and philosophy*, 1–45.

Kennedy, C., & McNally, L. (2010). Color, context, and compositionality. *Synthese*, 79–98.

Lassiter, D. (2009). Vagueness as probabilistic linguistic knowledge. In *International workshop on vagueness in communication.*

Leffel, T., Xiang, M., & Kennedy, C. (2016). Imprecision is pragmatic: Evidence from referential processing. In (pp. 836–854).

Persaud, K., & Hemmer, P. (2016). The dynamics of fidelity over the time course of long-term memory. *Cognitive Psychology*, 1–21.

Potts, C., Lassiter, D., Levy, R., & Frank, M. C. (2016). Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics*, 755–802.

Quine, W. V. (1969). Word and object. *Cambridge, Mass*.

Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PLOS ONE*.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, 1436–1441.

Schmidt, L. A., Goodman, N. D., Barner, D., & Tenenbaum, J. B. (2009). How tall is tall? compositionality, statistics, and gradable adjectives. In *Proceedings of the 31st annual conference of the cognitive science society.*

Sims, C. R., Ma, Z., Allred, S. R., Lerch, R. A., & Flombaum, J. I. (2016). Exploring the cost function in color perception and memory: An information-theoretic model of categorical effects in color matching. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

Stickles, R. T., Elise. (2014). The relation of color naming and the environment. In *Cogsci.*

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, 251–278.

Van Deemter, K. (2006). Generating referring expressions that involve gradable properties. *Comput. Linguist.*, 195–222.

Van Deemter, K. (2012). *Not exactly: In praise of vagueness*. Oxford University Press.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, 348.