The elusive benefits of vagueness

Dan Lassiter

Stanford University

April 18, 2018

This is a very interesting paper on an important topic. My main critical comment would be that it is unclear what the inclusion of experiment 1 adds to the paper. The experiment seems to have been problematic, as described in the beginning of the presentation of experiment 2. As a result of this skepticism about the cause of the results I found the attempts to interpret the results theoretically in the discussion to be unconvincing. The results from the second experiment seem to be much more solid and theoretically important. Perhaps the paper could be revised to present and interpret the second experiment only, treating the first as a pilot that revealed some issues that were resolved by the main experiment?

A second point is that the discussion of experiment 2 presents some results as more unequivocal than they appear to be. I have some detailed comments below on this.

A third question is whether the context-dependence of many vague expressions could be considered as a source of utility. E.g., *tall* and *warm* can adjust their meanings in ways that allow them to communicate more information about the location of some object relative to the statistical distribution of objects in a reference class. *1.83 meters tall* and *94C* don't have this ability. I've often thought that this flexibility is one of the main sources of utility for vague expressions with this property, since it allows our lexicon to be fairly expressive without having an enormous number of items.

Small comments

- While I see the value of presenting Lipman's example as such, the example is fairly strange for the purpose he wants to put it to, since it's not at all obvious that in the context describe *187.9 cm* would indeed be more conversationally useful than *tall* as a description of a person's height. I for one would have no idea how to pick out a person of that exact height from a crowd. Perhaps there is a more convincing modification of the example that would make the point, that we are sometimes inclined to use vague language even though a precise description would be uncontroversially more useful to the hearer?

- This sort of comes out in the discussion, but the experiment by Mishra et al. reported in section 2 seem to be of limited relevant since they confuse vagueness with lack of specificity. While *vague* is a typical way to describe an unspecific claim like *I was out with someone* as a response to *Where were you?*, this isn't the way we usually use the term in semantic analysis. So, this is an interesting experiment but I wonder if it is really worth discussing here.

- The Peters et al. experiment: the comment *it becomes doubtful whether any borderline cases could be conceived to arise in fact* – indeed it's not obvious but I would be surprised if they didn't. When given natural-sounding descriptions like *poor, fair, good, excellent* alongside hard-to-understand numerical values, I would guess that most participants would simply ignore the numerical values since they make less natural descriptions.

- Perhaps the data and analysis could be moved to a repo (e.g., github)? This way the URL remains stable even if someone moves institutions, if the university decides to change web addresses, etc.

- Could you include a sample display for each experiment presented?

- Why did the statistical analysis include only random slopes and not random intercepts? It would be standard to use both.

- p.13: I wouldn't describe indefinites like *a square with ...* as referring expressions

- exp 2 questions re presentation:
  - why was the central number always presenting in the middle, instead of fully randomizing?

- – *about 10* seems like a terrible description for both 6 and 15. My guess is this would cause subjects to boggle even if they eventually manage to make a guess.

- – what was the vague numeric instruction used in stimuli with larger dot arrays, e.g. (16,25,34)? [I assume it wasn't *about 10*]

- – for the numeric condition, the significant main effect reported at the bottom of p.15 (testing H1) is misleading, since the effect seems to be driven entirely by the much lower response time for crisp *6*. Everything else is nearly indistinguishable. So, it seems like the main thing to explain here is why crisp *6* is so much faster than other numerical expressions – presumably because it's a small number and there's no additional cost of trying to resolve the intention behind the fairly poor description *about 10* for 6?

- – perhaps the big main effect for instruction format: it seems like the interpretation of this result should be that numbers are in general harder to process than vague verbal expressions. This would seem to account for most of the variance in the response time data of experiment 2.

- – could the differences in the numeric vs verbal borderline response distribution shown in the right panel of fig. 2 be explained by stimulus design – i.e., the hypothesis I mentioned earlier, that *about 6* is just a poor description of both 6 and 15, with the result that participants were choosing more or less random with a slight bias for the number a little bit closer to 10? This would explain why there is almost no difference between borderline and expected in the vague/numeric condition.

- – p.17 *we found a small (but statisticslly significant) disadvantage of vague instructions* – again this is misleading since the effect is apparently driven almost entirely by a single condition (6, crisp) that was much faster than expected. There appear to be few differences otherwise . . . one option would be to exclude the 6:15:24 condition and then re-run the statistical analysis to see if any other effects emerge.

- – p.18 *allowed participants to use a comparison "algorithm"* – the work of Justin Halberda and collaborators on the verification strategies associated with *most* and *more than half* is highly relevant here (and indeed throughout the paper, given that *most* is vague but otherwise fairly similar in meaning to *more than half*)

- – p.19 Testing (H1), we have *Vague instructions were non-significantly easier* . . . do you mean *not significantly easier*? In any case from the graphs on the next page this seems misleading – what it looks like is a substantial main effect where vague instructions have a main effect of lowering the response time, but this effect is masked by interactions with quantity, most notably (again) the huge speed-up for 6 in the matching/crisp condition. I'd recommend following the method described in Roger Levy's paper on finding main effects in the presence of interactions, available here: `https://arxiv.org/abs/1405.2094`.

- – p.20 *the cost reduction account was wrong to predict significant main effect advantages for vagueness* – this seems way too strong given the issues I've just described. First, the statistical analysis needs to be redone as I just mentioned. Second, in the kind of statistical analysis being employed here it's not possible to show that something doesn't exist simply by failing to find evidence that it does exist. (APA guidelines: *Never use the unfortunate expression* accept the null hypothesis, cf. Wagenmakers, `http://www.ejwagenmakers.com/2007/pValueProblems.pdf`). At best, you can say *we failed to find evidence that* . . .

- – .p21, table 3: were the matching/crisp and matching/vague conditions unmatched in the target quantity as the table says? Or is this a typo?

- – p.22 *main interaction effect* - ??

- – p.22 *In the matching task vagueness slowed response times* – again this is not a good summary since there is a big interaction, and the apparent main efect may be driven by a single condition, here a slowdown for 6:15:24 in the matching/vague condition. Re-running using Levy's method would be advisable here.

- Typo: Paul Egré's name has an acute accent on the second *e*