


From: patrick.sturt@ed.ac.uk 
Subject: Language and Cognitive Processes - Decision on Manuscript ID PLCP-2011-OP-9118
Date: 21 February 2012 15:38:34 CET
To: mjglab@gmail.com, k.vdeemter@abdn.ac.uk
Cc: lcpadmin@csl.psychol.cam.ac.uk

21-Feb-2012

Dear Matthew,

I have now received three reviews of the paper that you submitted to Language and Cognitive Processes. As you will see from the reviews, which are attached at the bottom of this email, the reviewers raised a number of substantive concerns with the paper. On the basis of these reviews, I am afraid that I cannot accept the paper for publication in Language and Cognitive Processes.

As you will see, all three reviewers converge on a central problem with the paper, which Reviewer 2 summarises in his/her final comment: "I suspect that the main empirical contribution of these data are related to verification procedures rather than to game theoretic notions". From my own reading of the paper, and of the expert reviews, my impression is that you would need to carry out new experiments in order to make a good test of the theories that you are examining. Having said that, I do think that you are working on an important and interesting topic, and, as you point out in the introduction, the previous literature leaves several open questions that need to be addressed empirically, so I hope you will continue to follow this path.

I hope you will find the reviews to be constructive and helpful, and also that you will consider LCP as an outlet for your future research.

With best wishes,

Patrick Sturt

Associate Editor, Language and Cognitive Processes
patrick.sturt@ed.ac.uk

Reviewer(s)' Comments to Author:

Reviewer: 1
Comments to Author
Summary
=====

This paper reports on a series of three experiments investigating the effects of vagueness versus precision in referring expressions, with the goal shedding light on the prevalence of vague expressions in natural language. The authors seek to test the "cost reduction" hypothesis, namely that expressions with vague or loosely defined meanings are easier to process than those with crisp meanings. The domain of inquiry is quantification; cardinal numerals as representative of the crisp category are compared to the vague quantifiers 'many' and 'few' as representative for the vague category. The first two experiments appear to show an advantage for vagueness over precision; however, the third experiment is taken to show that respondents' behavior actually reflects an advantage for verbal vs. numerical format.

Review
=====

From the perspective of semantics and pragmatics (my field), the question of why vagueness exists is an important one, and I am particularly sympathetic to the use of experimental approaches to investigate potential advantages for vagueness over precision. In light of this, I find the results presented here very interesting, and potentially publishable. However, I have concerns that the experiments reported may not in fact test what they are purported to test, limiting what can be concluded about the advantages/disadvantages of vagueness or for that matter of verbal vs. numerical format. The authors to some extent acknowledge this, but the confounds are in my opinion more significant than what they describe.

The central issue as I see it is that the tasks in the vague vs. precise conditions in Exp. 1 and 2, as well as the verbal vs. numerical conditions in Exp. 3, lend themselves to different selection algorithms. The vague quantifier 'many' can be analyzed as meaning 'more than n, for some contextually determined n' (for few, replace 'more than' with 'less than'). In combination with the definite determiner 'the', which introduces presuppositions of existence and uniqueness, 'the square with many dots' necessarily refers to the square with the larger number of dots (*mutatis mutandi* for few). That is, in the vague condition in Exps. 1 and 2, the task can be performed simply by determining which of the two squares has more dots. In the precise condition, by contrast, the task involves determining which of the two squares matches the specified numerosity, which potentially requires considering each of the two squares separately. The authors make similar observations in their discussion of Exp. 2, but without directly confronting the possibility that the main effects identified could be entirely due to a difference between a "comparison algorithm" (which square has more dots?) and a "matching algorithm" (which square has n dots?), and not to the meaning of the referring expressions themselves.

Furthermore, the issue plausibly persists in Exp. 3. The indefinite description 'a square with many dots' does not presuppose that there is one and only one square that satisfies the description, but presumably respondents inferred that at least one of the three squares qualified (otherwise there is no acceptable answer). To avoid confronting the problem of possible borderline cases, a safe strategy for the respondent is to always choose the square with the largest number of dots, since if any of the three qualifies as 'many', it does. Thus in the Verbal Vague condition, just like the Verbal Precise condition, the task can be completed with a "comparison algorithm". The very similar response patterns in these two conditions tends to support that respondents in fact used the same procedure in both cases. As such, these findings do not allow conclusions about the existence of advantages/disadvantages for vagueness (at least for verbal quantifiers). More significantly, the numerical conditions again require a "matching algorithm", executed either precisely or vaguely.

It thus seems possible that the apparent advantage for vagueness found in Exps. 1 and 2, and the apparent advantage for verbal format found in

Exp. 3, really reflect an advantage for a comparison algorithm over a matching algorithm – it is easier to judge which array has more/the most dots than to judge which has, say, 10 dots. This is independent of vagueness per se, and of numerical vs. verbal format. From my understanding of the literature on number cognition (e.g. Dehaene 1997, Feigenson et al. 2004, Izard & Dehaene 2008 and much similar work), such an advantage is very plausible: a large body of research has shown that humans are very good at determining which of two arrays has more elements (as long as the ratio between their numerosity is sufficiently great), but not particularly good at quickly determining how many elements an array contains – particularly when the number is large, and the answer must be given precisely. (As a more general comment, I am somewhat surprised that the authors do not cite this literature, as both the central findings and the experimental paradigms are relevant to the present work.)

It is my recommendation that this potential confound be addressed before this work is published. Further empirical work seems necessary. Two possible follow-up experiments come to mind: 1) to eliminate the confound due to type of selection algorithm, one might compare ‘find the/a square with many dots’ to ‘find the/a square with more than 10 dots’, as both of these would be solvable with a comparison algorithm. My own intuition is that the ‘many’ condition would be easier (though admittedly this doesn’t separate vagueness from verbal format). I can’t at the moment see how one would tease apart vagueness vs. numerical/verbal format without introducing other confounds, but I wonder if this would require a different sort of task than a forced choice. 2) To test comparison vs. matching directly, one might eliminate the linguistic/numerical version of the specification as follows: first show one array of dots, then 2/3 arrays with the instruction ‘find the square with the same number of dots’ or ‘find the square with more dots’ (or even “=” vs “>”). Similar results to those found in Exp. 3 would tend to support that the difference is due to the type of algorithm employed, and not to character (vague/precise; numerical/verbal) of the referring expression.

With further follow-up experiments of this sort, this research could offer important insights into whether – and for what reason – vague expressions such as ‘many’ are advantaged in processing over precise ones.

I would also suggest a change in the presentation, namely to introduce at an earlier point in the paper the potential issues relating to testing the effects of vagueness independently of other factors. In the paper’s present form, the first two experiments are introduced as testing the effects of vagueness, and it is only later explained that they may not be able to do this; this is a little unsatisfying.

Additional Comments/Questions

=====

-If the cost reduction hypothesis relates to the hearers’ processing of vague vs. precise expressions, I wonder if it would be worth measuring reading time for the stimuli, in addition to response time in verification.

-In Exp. 1 it might have been preferable to classify stimuli according to the ratio of the two numbers rather than the numerical gap size, as the discriminability of two numbers/arrays is known to be ratio dependant (Dehaene 1997).

-For Exp. 1, the authors should clarify how they define subitizable. Judging from the classification of the test items it seems that 2 and 3 were considered subitizable while 4 was not; but I believe some authors (e.g. Trick & Pylyshyn 1994) would conclude that 4 falls within the subitizable range for adults.

-The treatment of borderline cases in Exp. 3 raises some questions. The authors seem to assume that out of 3 arrays, 1 will be considered a clear case of the vague quantifier, 1 a borderline case, and 1 a clear negative case, but this cannot be taken for granted (for example, respondents might instead judge there to be 2 clear cases). More specifically, for the 6/15/24 array, it is proposed that 6 is the best case of ‘about 10’ while 15 is borderline. However 6 and 15 are nearly equidistant from 10; furthermore, mental representation of number has been argued to be compressive (Dehaene 1997), and as such 15 might be perceived as actually closer to 10 than is 6. This compressive effect might also explain the finding that the ‘borderline’ option was chosen less often for the larger than the smaller quantity: when the target is the larger number both actual numerical distance and perceived distance would favor the largest of the 3 arrays (i.e. disfavor the ‘borderline’ response).

-The Number Vague vs. Number Crisp comparison in Exp. 3 is very interesting, in that it appears to show a (negative) effect of vagueness, without the potential confound discussed above. In this regard, though, it should be noted that some authors (e.g. Sauerland & Stateva 2007) would consider ‘about 10’ to be an instance of imprecision rather than vagueness. The authors might explicitly address whether or not results from expressions of this form can be generalized to more classically vague expressions such as ‘many’.

-Relatedly, in Fig. 6 it is seen that responses in the Number Crisp condition vary markedly with the numerical magnitude of the dot arrays. This seems potentially significant, but it is not addressed in the text

Minor point

=====

-In the abstract, it is stated that there were “diminishing returns for vagueness as the number of dots grew larger”. Should this be “...as the size of the gap became larger”?

References

=====

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford: Oxford University Press.

Feigenson, L., Dehaene, S. and Spelke, E. (2004). Core systems of number. *Trends in Cognitive Science* 8:307–314.

Izard, V. and Dehane, S. (2008). Calibrating the mental number line. *Cognition*, pp 1221-47.

Sauerland, U. and Stateva, P. (2007). Scalar vs. epistemic vagueness. In *Proceedings of Semantics and Linguistic Theory XVII*. Ithaca, NY: CLC Publications.

Trick, L., and Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 331–351.

Reviewer: 2

Comments to Author

This paper seeks to test predictions of game theory empirically, by contrasting precise and vague expressions in the context of numerical comparison.

I had two key difficulties with this paper that make me think that the paper may not be suitable for publication at this time.

First, the measures used to test the hypothesis at hand appear to depend critically not only on differences in meaning (and the corresponding game theoretic differences in information processing), but also in the verification procedures that are typical for numerals and quantifiers. The authors lay out their empirical program as follows, in the Introduction:

“Game theory and the cost reduction hypothesis both make claims about vagueness in terms of its effects on people engaged in communicative linguistic acts. For game theory, when vagueness influences communication, the term utility is used to capture this influence: vagueness is said to have less utility than crispness. The cost reduction hypothesis notes that we use vague language frequently, and assumes that the reason for the high frequency of use is that vagueness brings about an advantage, for producer or comprehender. In this paper, we set out to measure these effects so that we might have some empirical basis for preferring one account over the other. An obstacle to this effort is that whereas we can obtain, in the laboratory setting, several quantitative measures of the effects of language, the opposing game theory and cost reduction accounts do not specify a metric in which utility or advantage should be measured. Therefore it is for us as experimenters to suggest a way to capture and measure the effects of vague and crisp expressions such that they can meaningfully be compared. This requires us to choose, and motivate the choice of, both a task that brings about some measurable behaviour (or alternatively, some measurable physiological state) that is consequent upon whether the language used to bring it about was vague or crisp; and also a measure of the behaviour that is sufficiently fine-grained to distinguish the vague case from the crisp case.”

Having made this statement, the authors do not actually go about defining utility, advantage, or cost reduction, as they relate to linguistic processing. This is critical, because it is easy to conjure scenarios in which precise statements are more useful, in some sense, but nonetheless more difficult to verify and thus take longer to assess. For example, in a cash transaction between a buyer and a seller, precise language will be essential to both interlocutors, despite the fact that it may be more time consuming to count cash, rather than to estimate dollar amounts based on stacks of bills. A task that uses RT will find an advantage for vague language in this case, despite the fact that it seems obvious that precise language has greater utility by almost any conceivable definition that is sensible in this scenario. More generally, for any expression containing a numeral we can create a context in which we care about precise number. Whether vague or precise expressions have greater utility is not a question that can be answered in the abstract – it will always depend on context (otherwise, why would large numbers exist at all according to game theory?).

My second comment is that the individual experiments are consistent with just about any theoretical account. It's hard to imagine any theory that would not predict the results of Experiment 1, for example, given that we know that large numbers require counting when assessing precise cardinalities, but that vague quantifiers do not require counting. The paper would be much stronger if each study began with contrasting hypotheses and predictions. Currently, although hypotheses are sometimes mentioned, it's not clear that the authors are really engaging in hypothesis testing, since it's hard to imagine a theory that would not predict the results they find.

Until these steps are made, the paper amounts to a demonstration that quantifiers and numerals engage different verification procedures, and that for large numbers counting is more time consuming and clumsy, making numerals harder to use. But this result does not clearly speak to questions of informativeness, utility, advantage, etc, in the sense intended by game theory. It's not clear to me whether a simple reframing of the paper can address this issue. I suspect that it cannot. Instead, I suspect that the main empirical contribution of these data are related to verification procedures rather than to game theoretic notions, unless they can be defined in terms of verification in some way.

Reviewer: 3

Comments to Author

I feel that the main problem with this paper is that it has the wrong focus. The experimental results are not that interesting from the perspective of vagueness, but are quite interesting from the perspective of numeral versus quasi-numeral modification. I will begin this review with a short summary (hopefully revealing how I understood and misunderstood the paper) before proceeding with the commentary.

Summary (to reveal my understanding of the paper)

This paper presents itself as testing hypotheses about why vague terms are used instead of more precise terms. In particular, it explores one hypothesis that vague terms are easier to process than more precise terms, even when the precise terms are more accurate. There are three experiments in the paper, two of which are almost identical in design.

Experiment 1 tested reaction time in choosing between two sets when given with instructions using numbers (e.g., can you pick the square with two/three/four/etc dots?) as opposed to vague modifiers/quantifiers (e.g., can you pick the square with many/few dots?). The results as I understand them...

- (1) When asked to choose the square with fewer dots, the subjects did so faster with numerical instructions as opposed to vague ones.
- (2) When asked to choose the square with more dots, the subjects did so faster with vague instructions as opposed to numerical ones.
- (3) The enhanced speed with numeral was only significant when the set being chosen was subitizability.

Experiment 2 was identical to experiment 1 except that all the set-numbers used in the stimuli were not subitizable. The results as I understand them...

- (1) Vague instructions resulted in faster reaction times than numerical ones, the speed of reaction time was greater the closer the sets were in cardinality. The reaction time for pairs of sets that were easily discriminable in terms of cardinality did not demonstrate any advantage for vague vs. precise instructions.
- (2) The affect of numerical instructions speeding up the choice of smaller sets disappeared.

Experiment 3 involved reaction times in choosing between 3 sets. There were four different types of constructions: one with precise numerals (e.g., can you pick the square with 10 dots?), one with vague numerals (e.g., can you pick the square with about 10 dots?), precise modifier/quantifier (e.g., can you pick the square with the fewest/most dots?), vague modifier/quantifier (e.g., can you pick the square with few/many dots?). The

results as I understand them...

(1) The use of numerical expression, whether vague or precise, resulted in slower reaction times.

Main Comments:

I have no issue with the design of the experiment or the statistical analysis of the results. Most of my comments pertain to the discussion of the results and the context of the experiment.

(1) I am not sure why this is framed as a paper about vagueness. As experiment 3 demonstrates, the experimental design is not testing vagueness at all. Rather it is testing the difference between processing numerical versus quasi-numerical quantifiers (e.g., 10 vs. many). If the authors are attempting to appeal to those interested in vagueness, they have fail in that their experiments, by their own admission, do not test vagueness. If they are attempting to appeal to those interested in the processing of number words, for which there is a large audience, then they should refocus their paper from the beginning to speak about expected differences between number words and quasi-numerical quantifiers like "many" and "few." Vagueness is only one property that distinguishes these types of modifiers/quantifiers. For example, there are the following properties:

(A) Unlike number words, "few" and "many" cannot combine with explicit estimate terms such as "about" and "approximately" (e.g., about ten, approximately ten, *about few, *approximately few).

(B) Unlike number words, "few" and "many" are sensitive to comparison classes. For example, in sentences like "Many of the countries that start with the letter P are not democratic", what counts as "many" depends on how many countries there are that start with P. If there are only 7 countries that do so, then 4 would count as many, however if there are more than 200, then 4 would not count as many. Sentences like "Four of the countries that start with the letter P are democratic" are not sensitive in the same way. The truth conditions for the sentence remains the same no matter how many countries start with the letter P.

(C) Unlike number words, "few" and "many" can combine with superlative and comparative morphology. The word "few" can combine with -er (e.g., fewer) and -est (e.g., fewest). It has been argued that "most" is "many+est" and that "more" is "many+er" (see, Hackl 2009 or Klein, 1981, for example). The word "ten" on the other-hand cannot combine with such morphology (e.g., *tener, *tenest). These morphemes usually only combine with gradable modifiers, some of which are sensitive to comparison classes (e.g., tall) and some of which are not (e.g., wet).

Of these three properties, only (A) seems to have any correlation to vagueness. Adjectives like "wet" are not comparison class sensitive, are not vague, and yet combine with comparative morphology. Adjective like "tall" are comparison class sensitive, are vague, and combine with comparative morphology. Nouns like "heap" are not comparison class sensitive, are vague, and cannot combine with comparative morphology. Adjectives like "bald" are not comparison class sensitive, are vague, and yet combine with comparative morphology.

If the authors attempted to rewrite the paper as one that focuses on the differences between number words and quasi-numeral quantifiers, then they should address all the potential interesting differences that might play a role in processing.

Hackl M. (2009). On the grammar and processing of proportional quantifiers: most versus more than half. *Natural Language Semantics*, 17, 63-98.

(2) I wish that the authors attempted to discuss more the role of evaluation procedures that are connected to number words and quasi-numeral modifiers as it relates to their experiments. It seems obvious that even when using precise numbers in reference to non-subitizable quantities that speakers must engage in estimating exact cardinalities. In other words, to figure out which set has "10" requires estimating the cardinality of the two (or three) sets and choosing the one that is closest to 10. The same procedure (one assumes) would be used when evaluating terms like "about 10": estimating the exact cardinalities of each set and choosing the one closest to 10. Since counting is not an option given the task, the evaluation procedures would be equivalent. In light of this, it seems reasonable that one would not find any difference in processing time between the two types of instructions.

Note, however, that the evaluation procedure of "many" need not be the same. One need only order the sets from smallest to largest to figure out which one might count as many (largest one is guaranteed to be many, whereas the smallest one is guaranteed to be few). One does not need to engage in estimating cardinalities in the same way that the number words would require. In fact, space occupied by the dots could serve as a suitable method for establishing a relative order.

(3) Did the authors control for space as an estimate of cardinality. If so, they did not adequately high-light this aspect of their experiments. If not, then it would make an interesting addition to their current experiments. What if the size of the individual dots varied such that the set of 10 occupied the same amount of space as the set of 20 for example. Thus, subjects could not use space as an estimate of which set was bigger. They would be forced to estimate cardinality in order to obtain a ranking of sets. This seems relevant since the evaluation, and indeed the truth conditions, of words like "few" and "many" seem to depend on a ranking of sets in much the same way that "tall" depends on a ranking of individuals.

Unclear:

The difference between a symmetric versus asymmetric interpretation of vague terms was not adequately explained. Also, I think they inappropriately characterize the middle number as always being the borderline case. Although it is the borderline case for many and few, which function like gradable adjectives in that the borderline cases hover between a cut-off point in the middle of a ranking, it is not clear that "about 10" has one borderline case. Numbers above 10 and below 10 are both examples of borderline cases. So in fact there are two types of borderline cases for "about 10". This is the true division of the mean to characterize with their terms. It would be much better if they just explained it in more detail instead of using opaque terminology.