Kees van Deemter and Matthew Green

Vagueness & Rationality

The Elusive Benefits of Vagueness: the Evidence So Far

September 1, 2018

Contents

Chapter 1
The Elusive Benefits of Vagueness: Evidence from Experiments

Much of everyday language is vague, even in situations where vagueness could have been avoided (i.e., where vagueness is used "strategically"). Yet the benefits of vagueness for hearers and readers are proving to be elusive. We discuss a range of earlier controlled experiments with human participants, and we report on a new series of experiments that we conducted in recent years. These experiments, which focus on vague expressions that are part of referential noun phrases, aim to separate the utility of vagueness (as defined by the existence of borderline cases) from the utility of other factors that tend to co-occur with vagueness. Having presented the evidence, we argue that the evidence supports a view where the benefits that vague terms exert are due to other influences, and not to vagueness itself.

1.1 Introduction

In most academic use, the word 'vagueness' has a specific meaning. Keefe and Smith, for example, state "vague predicates have borderline cases, have fuzzy boundaries, and are susceptible to sorites paradoxes" (Keefe and Smith, 1997, p. 4), also Egré and Klinedinst (2011). The crucial criterion is the existence of borderline cases: "a word is precise if it describes a well-defined set of objects. By contrast, a word is vague if it is not precise" Lipman (2009, p. 1). A typical example is the word "tall", as applied to people for example, because here is no precise, known height which separates those who are tall from those who are not. The crucial point is that "tall" admits borderline cases (i.e., people who may or may not count as tall), which are the hallmark of vagueness as we use the term.

Linguists, philosophers of language, and more recently game theorists, have asked why natural languages contain so many vague expressions Lipman (2000, 2009), which are often used even in situations where the speaker could have used an expression that is not vague (i.e., crisp); in these situations we say that vagueness is used *strategically*. By introducing borderline cases, these expressions create potential misunderstandings, thereby creating "a worldwide several-thousand year

efficiency loss" Lipman (2009, p. 1). Lipman explains the point by means of a scenario in which a speaker describes a person to a hearer, who needs to identify that person in the arrivals hall of of an airport. In such a scenario, a precise description of the person's height (e.g., "The person's height is 187.96 cm") would be more useful than a vague one ("The person is tall"). Lipman uses this scenario to explain why standard game theory models of communication (e.g., Crawford and Sobel, 1982) predict that, under certain conditions, a crisp act of communication will always have more utility than a vague act that communicates the same state of affairs.

Lipman argued that the efficiency loss resulting from vague expressions would be unlikely to have arisen unless there are advantages as well as disadvantages associated with vague expressions. Lipman asked, essentially, what these advantages might be. Several tentative answers to Lipman's question have been offered (see van Deemter, 2009, 2010). One of the most promising answers appears to be the idea that vague expressions are easier to process, by a speaker and/or a hearer, than expressions that are not vague (i.e., crisp) (e.g., Lipman, 2009; De Jaegher, 2003; van Rooij, 2003). For example, Lipman writes: "For the listener, information which is too specific may require more effort to analyze" (2009, p. 11). We shall refer to this as the *cost reduction* hypothesis.

This article brings an experimental approach to these issues, focussing on vagueness in descriptions (e.g., "the square with few dots") and its effect on the hearer's ability to act on a given description, as measured by the time that it takes hearers to click on the referent of a description[1]. We find that, although hearer benefits from vague descriptions are straightforward to demonstrate in many cases, a closer experimental analysis militates against the conclusion that vagueness itself – as defined above, in terms of the existence of borderline cases – lies at the heart of these results. Instead, it is other factors, such as the presence of an overt numerical expression in the description, that proved to be decisive. We believe that, despite the fact that our experiments are unavoidably focussed only on a specific class of vague expressions (since any experiment can only deal with a limited number of different stimuli), these findings are potentially important, because they call into question whether "strategic" vagueness (i.e., vagueness where the speaker had a choice, because she could have been produced a crisp expression instead) has any advantages at all. In other words, returning to Lipman's question, it is possible that vagueness has evolved partly as a necessary evil (e.g., because of the limits of observation and prediction) and partly as a side effect of other factors.

---

[1] Other metrics could have been chosen, such as hearers' ability to remember information, for example, or error rates. Although error rates play a minor role in the present paper, for reasons that will become clear, we focus on response times in particular.

1.2 Related work

Charting the utility of vagueness is the attested aim of a small number of studies, but most of these have focussed on vagueness in a different sense, and focussing on different types of benefits for hearers. Two recent studies can illustrate both issues.

In a series of studies of behaviour modification, Mishra et al manipulated the presentation format of information about quantities in the domains of mental acuity, physical strength, and weight loss. In the weight loss study, participants were told that the study was designed to test the validity of a new (actually fictitious) health index, the HHI (Holistic Health Index). They were told that an ideal HHI score lies in the range of 45 to 55. In a longitudinal study, participants submitted their weight to a computer each week. Participants were told that two algorithms would be used to compute their HHI, and that the two might give different values initially, in which case the true score lay between the two values. In one condition, which the authors called the precise condition, the two algorithms gave the same score. In the other condition, which the authors called the vague condition, one algorithm added 3% to the score while the other algorithm subtracted 3% from the score, yielding a range of values whose midpoint was the same as the two values given in the precise condition.

One group of participants was given HHI scores in the ideal range: for this group their weight loss did not differ depending on whether they were given vague or precise HHI values. However for the other group, who were given HHI scores outside the ideal range, their weight loss was significantly greater if they were given vague HHI scores than if they were given precise HHI scores. The authors explain the improvement in the vague condition for this group as resulting from the participants' freedom to think of themselves a positioned on one end of the range - the end closest to the ideal HHI scores. This "illusion of proximity" (Mishra et al, 2011, p. 4) to the goal is argued to allow participants to generate positive expectancies that lead to behaviours that improve performance. In contrast, in the precise conditions, participants did not have this freedom of interpretation, and could not distort the information to bring about the beneficial *illusion of proximity*. These results are interesting, and of obvious potential practical importance. We note, however, that information presented as an exact range of values does not conform with the standard definition of vagueness (Keefe and Smith, 1997; Egré and Klinedinst, 2011), since an exact range does not admit borderline cases. In the terminology of Hobbs, the difference between a range and a single midpoint value is a difference of *granularity*. Furthermore, the experiments of Mishra et al did not explore benefits in terms of processing cost, but in terms of long-term behaviour change.

Similar issues arise from the work of Peters et al (2009). The authors carried out a series of studies where participants were required to rate hospitals based on various sources of information about quality of care. There was a between-subjects manipulation based on numeracy. The format of the information was manipulated within subjects: either numbers only were presented, or both numbers and evaluative categories were presented (e.g., *Poor*, *Fair*, *Good*, *Excellent*, with crisp visual boundary lines between the categories). Results showed that, for low-numeracy participants,

the presence of evaluative categories resulted in a diminished influence of an irrelevant affective state on the ratings. For all participants, the presence of evaluative categories resulted in better decisions and in a greater use of the most important and reliable types of information, such as survival rates.

It is, however, questionable whether the "evaluative categories" manipulation in this study can be considered a manipulation of vagueness. Certainly, terms like *Fair* admit the possibility of borderline cases. However, given that the boundaries between the categories were marked crisply, and that therefore the categories mapped crisply to numerical values, it becomes doubtful whether any borderline cases could be conceived to arise in fact. For example, *Fair* was mapped to 60% – 70% for the variable *percentage of heart attack patients given recommended treatment (ACE inhibitor)*. Accordingly, rather than the vagueness of categories such as *Poor*, Peters et al. emphasise the evaluative content inherent in these categories, and the affective potential of the evaluative content rather than the vagueness of the terms like *Fair*.

1.3 Our Approach to the Problem

Are vague expressions processed more easily by readers than crisp ones? Like Lipman, we focus on situations where numerical information is used in order to identify a referent. Reference, in other words, will be the communicative task on which we focus, partly because of the interest that this topic has recently drawn in various areas of Cognitive Science (van Deemter, 2016). By looking at one specific type of vagueness, we will be able to investigate the costs and benefits of vagueness relatively thoroughly. Whether our findings generalise to other uses of vagueness is a question on which we will speculate in the final section of this chapter.

We have chosen a narrative strategy in which we address a sequence of four experiments with human reader chronologically, explaining how each experiment helped us refine our research question. In order to do justice to our findings, we need to describe these experiments in a fair amount of detail.

Let us start by explaining the task that was given to the participants in our experiments. We used a *speeded forced choice* task to compare the processing costs of different references to quantities. In this context, speed and accuracy of responses are the key dimensions on which the different references can be compared. Each stimulus in the experiments was a set of dot arrays containing various number of dots, together with a preceding instruction (in the form of a referring expression) to choose one of the arrays with respect to its cardinality. The participant was asked to respond as quickly as possible while avoiding errors. We manipulated the instructions and the arrays in several ways across the four experiments.

All the experiments shared the following properties: Stimuli were created using the language GNU Octave (Eaton, 2002) and the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al, 2007). The position of the dots was randomised per-trial. The order in which trials were presented was randomised per-participant. There were 256 trials, presented in 4 blocks of 64 each, between which

the participant could rest. A MacBook Pro laptop computer with a 13 inch screen presented the stimuli to the participants and recorded responses. Participants were recruited using email lists at the University of Aberdeen, and paid ten pounds for participating. All participants self-reported fluency in English, and had normal, or corrected-to-normal vision. The experiment was conducted in a quiet room. Participants were asked to respond as quickly as possible while avoiding errors. There was a block of practice trials after which participants could ask any questions, following which the experimenter left the room. All *p* values reported for linear models were calculated using the R package *lmerTest* (Kuznetsova et al, 2016). The complete data and analysis for this experiment and the others in this paper are available at `https://github.com/mjgreen/vagueness`.

When the distance grows between two numbers, they become more easily distinguishable: the *numerical distance effect* has been shown for comparing the cardinality of two sets of dots (van Oeffelen and Vos, 1982) and for processing Arabic numerals and number words (Dehaene, 1996). We manipulated the number of dots in each array such that some sets of arrays had smaller numerical distances and others had larger numerical distances. Where a number was mentioned in the instructions, it was always in the form of an Arabic numeral. When two numbers are presented with the smaller on the left, this left-side presentation facilitates responses indicating the smaller number: the *Spatial-Numerical Association of Response Codes (SNARC)* effect (Dehaene et al, 1993; Gevers et al, 2006). We controlled which side the smaller number appeared on to avoid systematic influences of this effect.

There is abundant evidence (e.g., Trick and Pylyshyn, 1994) that very small (i.e., *subitizable*) quantities are recognised and processed by a distinct psychological mechanism that differs from that used to process larger quantities. We performed a pilot experiment (Green and van Deemter, 2011) in which we were able to confirm this finding in the experimental settings on which we are focussing in this paper. We found that, when participants were confronted with a stimulus consisting of two squares containing different numbers of dots[2], instructions of the form *Choose the square with* n *dots* led to consistently faster response times than instructions of the form *Choose the square with many/few dots* when $2 \leq n \leq 5$; the converse was true for $n > 5$. Given these findings, we henceforth focussed our studies on non-subitizable numbers, because it is there that vagueness is expected to have benefits.

In a second pilot experiment (Green and van Deemter, 2013), we again presented two dot arrays with the instruction in the form *Choose the square with ... dots*. The arrays contained larger numbers of dots than in the first pilot: one array always contained 25 dots, and the other contained either 5, 10, 15, 20, 30, 35, 40, or 45 dots. Each stimulus can therefore be seen in terms of the numerical difference between the number of dots in one array and the number of dots in the other: giving numerical distances of 5, 10, 15, and 20, with smaller numerical distances resulting in less discriminable arrays and larger distances resulting in more discriminable arrays. Our main manipulation was of the vagueness of the instruction, with two levels, *crisp* and *vague*. Assuming the dot array [5, 25], and the instruction refer-

---

[2] Such a stimulus is referred to hereafter as consisting of a set of *dot arrays*. The number of dots in an array is referred to as its cardinality. The physical arrangement of dots in each array is irregular.

ring to the smaller cardinality, the *crisp* instruction was *Choose the square with 5 dots* and its *vague* counterpart was *Choose the square with few dots*. We found, as expected, that responses were faster and more accurate for vague instructions than for crisp instructions. We also found an interaction between vagueness and numerical distance such that there were diminishing returns for vagueness as numerical distance increased, until, at the biggest numerical distance, there was no real difference any more between crisp and vague instructions. We interpreted this pattern as showing that cognitive load is relatively easy in both conditions when the stimuli are most discriminable, and so vagueness confers no additional advantage in those most easily discriminable stimuli.

However, the picture painted by these findings from the second pilot experiment might be misleading. First of all, there is a possibly confounding factor. Contrast an expression from the vague condition: 'the square with few dots' with an expression from the crisp condition: 'the square with 5 dots'. One difference is that 'few' has the potential for vagueness, whereas '5' is crisp. But another difference is that 'few' is verbal while '5' is numerical, in the sense that a number is mentioned explicitly. Since these two differences could not be separated in that experiment, the finding of a vagueness advantage is vulnerable to an alternative interpretation, that the crisp-vague difference was really an advantage for the verbal form of the quantifier. In our next experiment (reported below as Experiment 1) we therefore created verbal and numeric versions of each of the vague and crisp instructions so that we could compare the crisp-vague difference and the numeric-verbal difference in the same experiment.

In Experiment 1, reported below, we also addressed another issue with the second pilot experiment: participants chose one of two dot arrays - therefore the 'vague' quantifiers (few and many) uniquely identified one square. Recall our definition of vague – "a word is precise if it describes a well-defined set of objects. By contrast, a word is vague if it is not precise". The quantifiers few and many might not have realised their potential for vagueness in the context of a choice between (only) two alternatives - where there is no borderline alternative that could result in few and many being "not well-defined". Furthermore using definite articles in the instructions may have reinforced the participant's impression that only one choice counted as correct, and this impression could have been reinforced by the use of error feedback.

In the experiments we report below (Experiments 1, 2 and 3) we used what we learned from the pilot experiments to design situations that we believe did address the difference between crisp and vague instructions while taking account of the most important alternative explanations of this crisp-vague difference.

1.4 Experiment 1: separating vagueness from instruction format

To find out what happens when words are used in a context where their potential for vagueness comes to the fore, Experiment 1 used three arrays (rather than two) so

that the vague description had more than one possible referent, and used indefinite articles to avoid the impression that only one response counted as correct, and was carried out without error feedback. An indication that the potential for vagueness was indeed realised in Experiment 1 is that the borderline response was chosen fairly often: 16% of the time.

In Experiment 1, an item was an instruction followed by a set of three dot arrays defined by a triple of numbers, representing the number of dots in the left, middle, and right arrays. We used four different triples of numbers: (6,15,24); (16,25,34); (26,35,44); (36,45,54). Each set of arrays comprised three arrays (instead of two as in Green and van Deemter (2013)); the array representing the central number was always presented in the middle of the three; there were two flanking arrays where one had fewer dots than the central array and the other had more, and these flanking arrays appeared equally often on the left and right of the central array.

Table 1.1 gives the full set of stimuli and associated instructions. The way in which borderline responses were construed is as follows, using as an example the array (6:15:24) and instructions that identified the smaller flanking array (6). 6 was classified as the expected response. 15 was classified as the borderline response. 24 was classified as the extreme response.

- In the vague numerical condition the instruction was "Choose a square with about 10 dots" – none of the arrays contained exactly 10 dots, but 10 is closer to 6 than it is to 15, making 6 a better response to that instruction, 15 a borderline response, and 24 an extreme response.
- In the vague verbal condition we used "Choose a square with few dots". We considered this to be equivalent in terms of which responses were expected (6), borderline (15) and extreme (24).
- In the crisp numerical condition we used "Choose the square with 6 dots". The smaller flanking array always contained exactly the specified number of dots. We considered this to be equivalent in terms of which responses were expected (6), borderline (15) and extreme (24).
- For crisp verbal, we used "Choose the square with the fewest dots". We considered this to be equivalent in terms of which responses were expected (6), borderline (15) and extreme (24).

On each trial, first the referring expression that constituted the instruction for that trial was displayed (e.g., "Choose a square with about 10 dots"). Participants then pressed a key to indicate that they had read the instruction. The instruction remained on screen, and after 1000 ms, the arrays appeared. An example stimulus is given in Figure 1.1. Response time was measured from the presentation of the arrays until the keypress indicating the participant's choice. The trial would timeout after 60 seconds if there was no response. In this experiment, no feedback was given. This was because, in the vague conditions, we did not regard any response as "correct" or "incorrect", but instead as "expected response"; "borderline response"; and "extreme response", and we did not want to draw participants' attention to this distinction explicitly. Which choice the participant made was recorded for analysis.

Table 1.1 Experiment 1 instructions arranged by condition. The instructions given in the table started with "Choose . . . "

Item	Quantity	Number	Crisp	Vague
06:15:24	Small	Numeric	the square with 6 dots	a square with about 10 dots
06:15:24	Small	Verbal	the square with the fewest dots	a square with few dots
06:15:24	Large	Numeric	the square with 24 dots	a square with about 20 dots
06:15:24	Large	Verbal	the square with the most dots	a square with many dots
16:25:34	Small	Numeric	the square with 16 dots	a square with about 20 dots
16:25:34	Small	Verbal	the square with the fewest dots	a square with few dots
16:25:34	Large	Numeric	the square with 34 dots	a square with about 30 dots
16:25:34	Large	Verbal	the square with the most dots	a square with many dots
26:35:44	Small	Numeric	the square with 26 dots	a square with about 30 dots
26:35:44	Small	Verbal	the square with the fewest dots	a square with few dots
26:35:44	Large	Numeric	the square with 44 dots	a square with about 40 dots
26:35:44	Large	Verbal	the square with the most dots	a square with many dots
36:45:54	Small	Numeric	the square with 36 dots	a square with about 40 dots
36:45:54	Small	Verbal	the square with the fewest dots	a square with few dots
36:45:54	Large	Numeric	the square with 54 dots	a square with about 50 dots
36:45:54	Large	Verbal	the square with the most dots	a square with many dots

Choose a square with about 10 dots
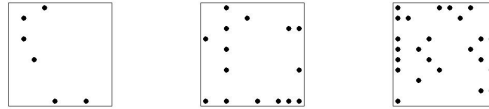
Fig. 1.1 Experiment 1 and 2 example stimulus

1.4.1 Hypotheses (Experiment 1)

We formulated the following hypotheses for Experiment 1:

Hypothesis 1 (Crisp/Vague RT)   Vague instructions should result in faster responses than crisp instructions; and this pattern should hold when the model is restricted to numeric-only data and when it is restricted to verbal-only data.

Hypothesis 2 (Numeric/Verbal RT)   There should be no real difference between responses to Numeric instructions and Verbal instructions (based on our interpre-

tation of the experiment in Green and van Deemter (2013), where we thought that vague instructions alone were driving the advantage for instructions that were both vague, and also in verbal format).

Hypothesis 3 (Item RT)   Responses should take longer as the number of dots in the display grows larger (i.e., as the levels of Item increase).

Hypothesis 4 (Response Type)   Vague instructions should lead to more borderline responses than crisp instructions.

1.4.2 Results (Experiment 1)

30 participants were recruited. Response times from all trials were trimmed at 2.5 standard deviations for each subject, leading to the loss of 236 trials, 3.1% of the data. The distribution of remaining response times was skewed with many long responses. These remaining response times were log-transformed which reduced this skew so that their distribution more closely approximated a normal distribution. Condition means for response times and proportion of borderline responses are given in Figure 1.2.
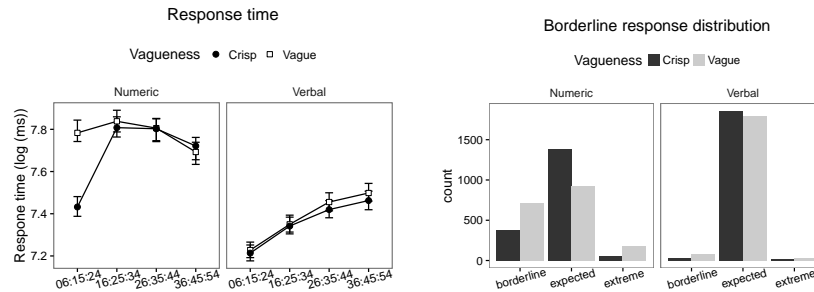

Fig. 1.2 Experiment 1 results: mean response times by condition and item, and counts of borderline case responses by condition.

A linear mixed model was constructed for the (log-transformed) response times, with sum-coded vagueness, instruction format, and their interaction, and item, as fixed effects, and per-participant intercepts and slopes for sum-coded vagueness, instruction format, and their interaction as random effects.

Test of Hypothesis 1 (Crisp/Vague RT)   Vague instructions actually led to significantly slower responses than crisp instructions, against Hypothesis 1: $\beta = 0.058$, $se = 0.013$, $t = 4.55$, $p < 0.001$. When the model was restricted to numeric-only instructions Vague instructions still led to significantly slower responses than crisp instructions $\beta = 0.093$, $se = 0.021$, $t = 4.51$, $p < 0.001$. When the model was restricted to verbal-only instructions Vague instructions tended to slow responses, but not significantly: $\beta = 0.024$, $se = 0.016$, $t = 1.46$, $p = 0.155$.

Test of Hypothesis 2 (Numeric/Verbal RT)   There was actually a significant difference between numeric and verbal instructions, with numeric instructions leading to longer responses than verbal instructions, against Hypothesis 2: $\beta 0.265 =$, $se = 0.072$, $t = 5.08$, $p < 0.001$.

Test of Hypothesis 3 (Item RT)   Responses took longer as the levels of Item increased, supporting Hypothesis 3: $\beta = 0.120$, $se = 0.017$, $t = 7.11$, $p < 0.001$.

However, given that the response time plot in the left panel of Figure 1.2 shows that responses to 6:15:24 in the "crisp numeric" instructions condition were extremely fast relative to the "vague numeric" instructions to 6:15:24, the effects in the model of the full dataset could be driven by this difference. A clearer picture of the effects of interest might be obtained by removing the 6:15:24 level of Item from the data set, and fitting the model to this restricted data. Doing this did not affect the direction of the effects in the full dataset, but whereas the effects were significant in the full dataset, they were not significant in the restricted dataset. Full details of the analysis of the restricted dataset are available in the online materials.

A generalized linear mixed model Jaeger (2008) was fit to the data for selection of the borderline response, with sum-coded vagueness, instruction format, (and their interaction), and item as fixed effects, and the same effects as slopes over participant, as well as per-participant intercepts as random effects. The distribution of responses over the nearest match square, the borderline square, and the furthest match square are given in Figure 1.2. Participants chose the borderline square on 16.6% of trials overall.

Hypothesis 4 (Response Type)   Participants were significantly more likely to choose the borderline option for vague instructions than for crisp instructions (21.9% vs 11.3%: $\beta = 0.62$, $se = 0.22$, $z = 2.8$, $p = 0.0059$). Participants were also significantly more likely to choose the borderline square when the instruction used the numerical format rather than the verbal format (30.1% vs 3.0%: $\beta = -3.35$, $se = 0.23$, $z = -14.6$, $p < 0.0001$).

1.4.3 Discussion (Experiment 1)

This experiment tested whether vague instructions would result in faster responses than crisp instructions, when borderline cases were present. Faster responses for vague instructions were found in pilot experiment B, but there were no borderline cases in that experiment.

In this experiment we found in contrast that vague instructions resulted in slower responses than crisp instructions: a difference that was significant when considering the full data (112ms), but which was not significant after removing the smallest arrays from the analysis, which had a pattern opposite to the main trends in the rest of the data.

We also found that the effect of instruction format was significant, with numerical format slowing responses by 689 ms on average, such that the disadvantage

of numerical format overwhelmed the contribution of vagueness. The verbal vague condition still yielded faster responses than the numerical crisp condition, so the pattern from pilot experiment B was reproduced, but in the light of the evidence from this experiment (Experiment 1), in the presence of borderline cases, the advantage that was ascribed to vagueness before now looks more like an advantage of verbal instruction format.

However, once again there is a possibly confounding factor. Observe that, in Experiment 1, instruction format (i.e., the difference between numeric and verbal) went hand in hand with might be called the (human) "selection algorithm": To see this, consider the task of selecting the dot array that contains "few dots":" to do this, it suffices to *compare* the three arrays and select the one that contains the fewest elements. To select the dot array that contains "16 dots" seems to require the participant to estimate, and then *match*, the cardinality of (at least) one dot array to 16, a process which could plausibly take longer, independently of vagueness. Therefore, our results so far permit the interpretation that what made the instructions in the verbal condition fast is not the fact that they were worded verbally, but that they allowed participants to use *comparison* rather than having to resort to *matching*.

In the next two experiments we pitted the comparison algorithm and matching algorithm selection tasks against each other while controlling vagueness and instruction format. In Experiment 2 we restricted all the instructions to numeric quantifiers while factorially manipulating vagueness and selection task. In Experiment 3 we ensured that all instructions used verbal quantifiers, while also factorially manipulating vagueness and selection task. This allowed us to distinguish between the predictions of the selection task account and the instruction format account.

1.5 Experiment 2: focus on instructions that contain numerals

The main aim of Experiment 2 was to see whether vagueness would exert beneficial effects when all conditions used numerals in the instructions, and when there were vague and crisp versions of the instructions for both comparison and matching strategies. The main changes from Experiment 1 were that the human selection task was explicitly controlled (i.e., whether the task amounted to matching or comparison), and that all conditions were constrained to mention a number. We used the same arrays as in Experiment 1 (an example stimulus is given in Figure 1.1). We used a 2 x 2 factorial manipulation of vagueness and selection task (see Table 1.2). On each trial an instruction was presented: participants pressed a key to dismiss the instruction, at which time the dot arrays were presented until the participant responded, and the response time and choice were recorded. Table 1.2 shows the instructions for each condition. Note the difference between "fewer than 20" and "far fewer than 20": whereas the former cannot have borderline cases (i.e., for each number it is clear whether the number is smaller than 20 or not), the latter can.

Table 1.2 Experiment 2: Instructions arranged by condition. The instructions given in the table started with "Choose a square with ..."

Item	Quantity	Selection	Crisp	Vague
06:15:24	Small	Comparison	fewer than 20 dots	far fewer than 20 dots
06:15:24	Small	Matching	6 dots	about 10 dots
06:15:24	Large	Comparison	more than 10 dots	far more than 10 dots
06:15:24	Large	Matching	24 dots	about 20 dots
16:25:34	Small	Comparison	fewer than 30 dots	far fewer than 30 dots
16:25:34	Small	Matching	16 dots	about 20 dots
16:25:34	Large	Comparison	more than 20 dots	far more than 20 dots
16:25:34	Large	Matching	34 dots	about 30 dots
26:35:44	Small	Comparison	fewer than 40 dots	far fewer than 40 dots
26:35:44	Small	Matching	26 dots	about 30 dots
26:35:44	Large	Comparison	more than 30 dots	far more than 30 dots
26:35:44	Large	Matching	44 dots	about 40 dots
36:45:54	Small	Comparison	fewer than 50 dots	far fewer than 50 dots
36:45:54	Small	Matching	36 dots	about 40 dots
36:45:54	Large	Comparison	more than 40 dots	far more than 40 dots
36:45:54	Large	Matching	54 dots	about 50 dots

1.5.1 Hypotheses (Experiment 2)

For Experiment 2, we formulated the following hypotheses:

Hypothesis 1 (Crisp/Vague RT)   Vague instructions should result in faster responses than crisp instructions.

Hypothesis 2 (Comparison/Matching RT)   Instructions that allow comparison should result in faster responses than instructions that necessitate matching.

Hypothesis 3 (Interaction)   The vagueness effect should differ according to whether the selection task is comparison or matching.

1.5.2 Results (Experiment 2)

38 participants were recruited. Response times from all trials were trimmed at 2.5 standard deviations for each subject, leading to the loss of 204 trials (2.8% of the trials). The distribution of remaining response times was skewed with many long responses. These remaining response times were log-transformed which reduced this skew so that their distribution more closely approximated a normal distribution. Condition means for response times are plotted in Figure 1.3.

A linear mixed model was constructed for the logged response times, with sum-coded vagueness, selection task, and their interaction, and item as fixed effects, and
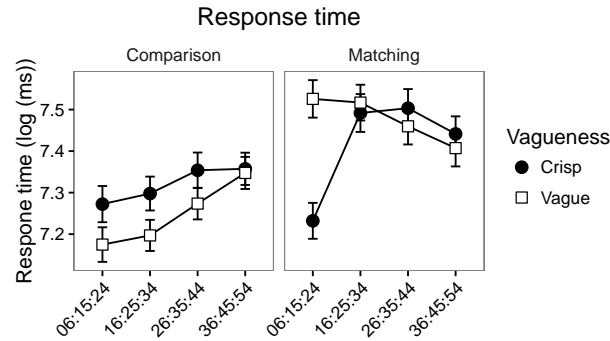
Fig. 1.3 Mean response times by condition for Experiment 2 where all instructions were numeric

per-participant intercepts and slopes for sum-coded vagueness, selection task, and their interaction, and item for random effects.

Test of Hypothesis 1 (Crisp/Vague RT)   Vague instructions resulted in faster responses than crisp instructions on average. However this difference was not significant in the full model ($\beta = -0.0057$, $se = 0.0137$, $t = -0.42$, $p = 0.678$). Using Levy's method (Levy, 2014) to test for main effects in the presence of higher-order interactions, by doing model comparison between a null model that included all interaction terms involving Vagueness but leaving out a term for the main effect of Vagueness, against a full model that differed only by including Vagueness as a main effect, showed that the full model was no better than the reduced model ($df = 1$, $p = 0.676$), consituting more evidence that Vagueness did not exert a significant main effect on response times.

Test of Hypothesis 2 (Comparison/Matching RT)   comparison instructions resulted in faster responses than matching instructions, and the difference was significant ($\beta = 0.1618$, $se = 0.0255$, $t = 6.34$, $p < 0.001$).

Test of Hypothesis 3 (Interaction)   Although Vagueness did not exert a significant main effect, Vagueness did exert effects in interactions with some other variables: the interaction between Vagueness and Selection task was significant ($\beta = 0.1306$, $se = 0.0205$, $t = 6.38$, $p < 0.001$), suggesting that Vagueness speeded RTs in the comparison condition but slowed them down in the matching task.

1.5.3 Discussion (Experiment 2)

The cost reduction account predicted that there should be a significant main effect of Vagueness such that responses would be faster for Vague intructions than for crisp instructions. We found that although there was a very small effect in that direction,

the effect was not statistically significant. Models that differed only in the presence of Vagueness as a main effect were shown not to differ significantly in their explanatory value. However we did find that Vagueness exerted effects on other variables: Vagueness speeded RTs in the comparison task and slowed RTs in the matching task.

1.6 Experiment 3: focus on instructions that do not contain numerals

This experiment mirrors Experiment 2, but focussing on instructions that did not use a number. We manipulated vagueness and the selection task (comparison and matching). In order to implement the experiment without mentioning numbers in the instructions, we changed the sequence of each trial to include a 'target' (i.e., a dot array of a particular cardinality) before the array, so that we could then refer to the target's cardinality in the instruction using expressions like *the same number of dots as the target*; *fewer dots than the target*. An example of this sequence is given in Figure 1.4. This presentation of a target before the main body of the trial shares some features with Izard and Dehaene, Experiment 2, although in that experiment participants were told the cardinality of the target (called an *inducer* in that paper) whereas in our experiment we did not tell participants the cardinality of the prime array. An item was thus a combination of a target dot array, an instruction that did not contain a number, and a set of dot arrays taking their cardinalities from the same triples used in Experiments 1 and 2. Table 1.3 spells out how the instructions were constrained not to mention a numeral and gives examples of targets.

Table 1.3 Experiment 3 instructions arranged by condition. The instructions given in the table started with "Choose a square with …"

Item	Target	Selection	Vagueness	Instruction
06:15:24	6	Matching	Crisp	the same number of dots as the target
06:15:24	10	Matching	Vague	about the same number of dots as the target
06:15:24	20	Comparison	Crisp	fewer dots than the target
06:15:24	20	Comparison	Vague	far fewer dots than the target

1.6.1 Hypotheses (Experiment 3)

For Experiment 3, we hypothesised:

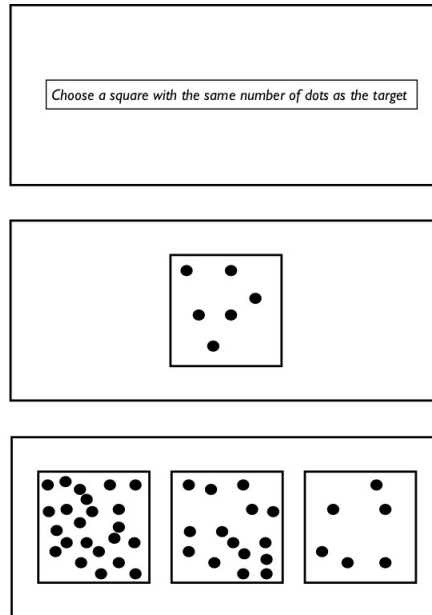*Choose a square with the same number of dots as the target*

Fig. 1.4 Experiment 3 example stimulus

Hypothesis 1 (Crisp/Vague RT)   Vague instructions are easier for the reader than crisp ones.

Hypothesis 2 (Comparison/Matching RT)   Comparison is easier for the reader than matching.

Hypothesis 3 (Interaction)   The effect of vagueness differs depending on whether the selection task mandated by the instructions is matching or comparison.

1.6.2 Results (Experiment 3)

40 volunteers participated. Response times from all trials were trimmed at 2.5 standard deviations for each subject, leading to the loss of 211 trials (2.8% of the trials). The distribution of remaining response times was skewed with many long responses. These remaining response times were log-transformed which reduced this skew so that their distribution more closely approximated a normal distribution. Condition means for response times are plotted in Figure 1.5.

A linear mixed model was constructed for the logged response times, with sum-coded vagueness, selection task, and their interaction, and item as fixed effects, and per-participant intercepts and slopes for sum-coded vagueness, selection task, and their interaction, and item for random effects.
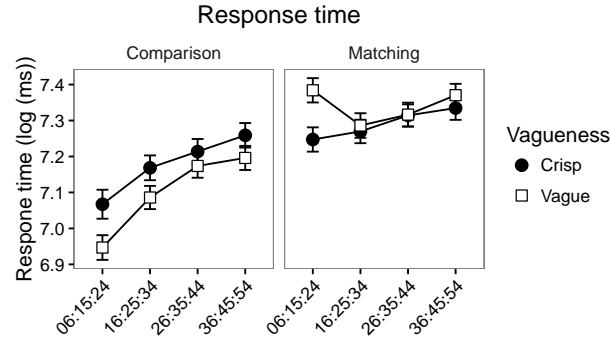
Fig. 1.5 Mean response times by condition for Experiment 3 where all instructions were verbal

Test of Hypothesis 1 (Crisp/Vague RT)   Vague instructions resulted in faster re-
sponses than crisp instructions on average. However this difference was not sig-
nificant in the full model ($\beta = -0.015$, $se = 0.0097$, $t = -1.58$, $p = 0.119$).
Using Levy's method (Levy, 2014) to test for main effects in the presence of
higher-order interactions, by doing model comparison between a null model that
included all interaction terms involving Vagueness but leaving out a term for the
main effect of Vagueness, against a full model that differed only by including
Vagueness as a main effect, showed that the full model was no better than the
reduced model ($df = 1$, $p = 0.118$), consituting more evidence that Vagueness
did not exert a significant main effect on response times.

Test of Hypothesis 2 (Comparison/Matching RT)   Comparison instructions resulted
in significantly faster responses than matching instructions ($\beta = 0.176$, $se = 0.0168$, $t = 10.51$, $p < 0.001$).

Test of Hypothesis 3 (Interaction)   Although Vagueness did not exert a significant
main effect, Vagueness did exert effects in its interaction with Selection task ($\beta = 0.123$, $se = 0.0166$, $t = 7.42$, $p < 0.001$), suggesting that Vagueness speeded
response times in the comparison condition but slowed response times in the
matching task.

1.6.3 Discussion (Experiment 3)

The cost reduction account predicted that there should be a significant main effect
of Vagueness such that responses would be faster for Vague intructions than for
crisp instructions. We found that although there was a very small effect in that di-
rection, the effect was not statistically significant. Models that differed only in the
presence of Vagueness as a main effect were shown not to differ significantly in their
explanatory value.

However, the results also showed that vagueness did exert an influence on reaction times according to whether the task was comparison or matching: vagueness was beneficial for comparison and detrimental for matching (the same as Experiment 2) even when no numbers were allowed in the instructions.

1.7 Discussion of experiments 2 and 3

The main aim of these two experiments was to test whether vagueness confers any cognitive benefits over and above those due to differences in the selection task according to whether the instruction mandates a *comparison* selection task or a *matching* selection task, when number-use is held constant. The main effect of selection task showed that the assumption that the *comparison* task is easier than the *matching* task is well-founded. In both experiments people were reliably quicker to respond in the *comparison* task.

Vagueness, which was the phenomenon on which our investigation focussed, did not exert a significant main effect in response time. However when the comparison and selection tasks were analysed separately, there was small significant advantage for vagueness in the *comparison* tasks, but a small significant disadvantage for vagueness in the *matching* tasks.

1.8 General Discussion

To summarise our findings, we asked why strategic vagueness is as frequent as it is and we decided to focus on what we see as the most promising explanation, namely that vague expressions are easy to process by speakers and hearers: the cost hypothesis, as we have called it. We decided to test this hypothesis by experimentally investigating whether vague descriptions are resolved by hearers more quickly than crisp ones. Although we were able to find some interesting (and statistically highly significant) effects, it appears to us that if our sequence of experiments is assessed as a whole, it cannot be seen as confirmation of the cost hypothesis.

To explain why this is, let us summarise our findings so far: Experiment 1 showed that number avoidance in the verbal format instructions is an important factor driving the faster response times in the task, and that vagueness does not have any additional benefit in either the verbal format instructions or the numerical format instructions. However, Experiment 1 did not distinguish benefits of number avoidance from benefits of the comparison selection task. In Experiments 2 and 3 we manipulated vagueness and the selection task, separately at each level of numerical format. Across the two experiments, we found that the comparison-task instructions attracted faster response times than the matching-task instructions. Within the two experiments we found that vagueness exerts benefits when the selection task is *comparison*, but not when the task is *matching*.

What is one entitled to conclude? Given that we were able to identify a class of situations in which vague expressions led to faster response times than crisp ones, would it be valid to conclude that we have finally discovered an advantage for vagueness that cannot be ascribed to some other factor? We believe the answer to this question is negative.

Table 1.4 Vagueness as range reduction: a summary of Experiments 2 and 3

selection task	vagueness	candidates	effect of vagueness
comparison	crisp	2	vagueness advantage
	vague	1	
matching	crisp	1	vagueness disadvantage
	vague	2	

To see why, consider Figures 1.3 and 1.5. Both figures depict four conditions, depending on whether the expression was crisp or vague, and depending on whether the referent could be identified using a comparison strategy or not. Two of the resulting four conditions result in an expression that can denote either of two referents; the other two conditions result in an expression that can only denote one referent, with the other possible referent being a marginal candidate at best.

To see why vagueness has opposite effects, depending on whether it is used in matching or comparison situations, consider the stimulus with (6,15,24) dots. Now compare 'Choose a square with 6 dots' with its vague counterpart 'Choose a square with about 10 dots': by adding the word 'about', we broaden the range of squares that the expression might be referring to. On the other hand, compare 'Choose a square with fewer than 20 dots' with 'Choose a square with far fewer than 20 dots': by adding the word 'far', we did not broaden the range of squares denotable by the expression: we narrow it down, because only some of the squares that have fewer dots may have *far* fewer dots.

The benefits of vagueness in the *comparison* task in experiments 3 and 4 could thus be explained as differences in the number of valid targets for the expression. This leads us to speculate that the benefit for vagueness here could be due to the vague expression foregrounding a particular valid target while the crisp expression carries with it the additional task of distinguishing between two alternative valid targets, something we propose to call a "range-reduction" benefit.

The observation that conditions with 1 candidate lead to shorter response times than conditions with 2 candidates is consistent with the range reduction hypothesis, but not with the idea that vagueness is beneficial. It appears, in other words, that shorter response times will only result from a vague expression if this expression leads to range reduction. Once again, it is not vagueness itself that has advantages but a phenomenon (namely range reduction) that is an automatic concomitant of vagueness in some types of situations.

Looking at the entire series of experiments, our findings suggest that the observed benefits of vague expressions may be due to factors other than vagueness: factors like avoiding numbers; permitting comparison tasks; and range reduction. The picture that is starting to emerge is subtle: on the one hand, in the situations that we have been studying vagueness is not intrinsically beneficial. On the other hand, vague expressions frequently possess other features that *are* beneficial, and these are what give us the incorrect impression that vagueness itself is beneficial. Vagueness may thus have acquired a reputation that it does not deserve. The answer to Lipman's question, of why vagueness permeates human language (see our Introduction), may lie in a different direction after all, possibly relating to benefits for the speaker rather than the hearer.

A comparison may clarify the logic of the situation. In recent years a number of studies, focussing on red wine, have suggested that alcohol, consumed in low doses, may have health benefits. An alternative explanation, however, asserts that it is not the alcohol in the wine that was beneficial, but antioxidants from grapes. If this alternative explanation is correct, then alcohol may not be healthy after all.

Our findings suggest a re-think of the questions on which much research on the utility of vagueness rests. Years of research on the logic of vagueness – giving rise to such techniques as Partial Logic Fine (1975, e.g.,), Probabilistic Logic Edgington (1997), and Fuzzy Logic Zadeh (1965) – have primed the research community to expect that some special utility of vagueness is an important part of the answer, but our findings call this expectation into question.

Although our own studies in this article have focussed on vagueness in descriptive Noun Phrases only, it seems to us plausible that vagueness plays a similar role in other linguistic constructs. For example, consider reports on air temperature. Given a numerical temperature measurement or prediction, we might word it as

(a) *27.2 degrees Celsius*, or
(b) *approximately 27 degrees*, or
(c) *above 25 degrees*, or
(d) *warm*,

among other candidate expressions. Which of these descriptions is most effective, for example as part of a weather report? If the linguistic literature is to be believed, then options (a) and (c) convey crisp information, whereas (b) and (d) are vague (i.e., they permit borderline cases). In the situations studied in our own experiments and the ones discussed in section 1, we found no evidence that vagueness is beneficial for hearers. Rather than asking whether a candidate expression is vague, other questions might shed more light on the choice, similar to the ones identified in our studies. These questions might focus on the amount of information that a given expression conveys (i.e., on granularity), on the avoidance of numbers, and on the use of evaluative terms. Let's see how this might pan out for the above examples from the weather domain.

First, the experiments by Mishra et al suggest that it is important how much information is conveyed by an expression, and their findings are echoed by our own thoughts about range reduction (following Experiments 2 and 3). In the case of (a)–

(d) above, it appears that (a) conveys the most detailed information (designating the smallest segment of the temperature scale), followed by (b), then (d), then (c) (e.g., 40 degrees is above 25, but at 40 Celsius the word "warm" is likely to give way to "hot" or "scorching"):

$a < b < d < c$

If these hunches are correct, then it seems to us that it is relatively unimportant whether a given expression is vague or crisp. Other factors seem more important; moreover, it may depend on the task and the audience which of a-d is preferred. For example, an *expert* may prefer to read expression (a), because it gives her the most detailed information on which to base her decisions. On the other hand, expression (d) ("warm") is shorter than the other three and avoids the use of numbers; our experiments suggest that this may make (d) more rapidly understood than its competitors; earlier experiments point in the same direction, given the evaluative nature of "warm" (see our section 1 and Peters et al), which is especially important if the hearer is unfamiliar with the metric used. These considerations suggest that *non-experts* might prefer expression (d).

One way to see why vagueness (as defined in our Introduction) may not benefit human communication is the following thought experiment. Suppose a group of speakers understand the word "warm" as vague, agreeing that temperatures above 26 count as warm, and temperatures below 24 do not count as warm, but considering temperatures between 24 and 26 as borderline cases. Now one day these speakers agree to sharpen up their definition deciding that, henceforth, "warm" means "> 25 degrees" (as in (c) above): this decision resolves the borderline cases, while everything else remains the same. It seems unlikely that this change in language use, from a vague meaning to a crisp one (i.e., one that has no borderline cases anymore), would lower the utility of the word. Our experimental findings, and the conclusions that we draw from them, are consistent with this idea.

References

Brainard DH (1997) The Psychophysics Toolbox. Spatial Vision 10:433–436

Crawford VP, Sobel J (1982) Strategic information transmission. Econometrica 50(6):1431–1451

De Jaegher K (2003) A Game–Theoretic Rationale for Vagueness. Linguistics and Philosophy 26:637–659

Dehaene S (1996) The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. Journal of Cognitive Neuroscience 8(1):47–68

Dehaene S, Bossini S, Giraux P (1993) The mental representation of parity and number magnitude. Journal of Experimental Psychology: General 122(3):371

Eaton JW (2002) GNU Octave Manual. Network Theory Limited

Edgington D (1997) Vagueness by Degrees. In: Keefe R, Smith P (eds) Vagueness: a Reader. A Bradford Book, The MIT Press, Cambridge, MA

Egré P, Klinedinst N (2011) Introduction: Vagueness and language use. In: Egré P, Klinedinst N (eds) Vagueness and Language Use, Palgrave

Fine K (1975) Vagueness, truth and logic. Synthese 30(3):265–300

Gevers W, Lammertyn J, Notebaert W, Verguts T, Fias W (2006) Automatic response activation of implicit spatial information: Evidence from the SNARC effect. Acta Psychologica 122(3):221–233

Green MJ, van Deemter K (2011) Vagueness as cost reduction: An empirical test. In: Proceedings of 'Production of Referring Expressions' workshop at 33rd Annual Meeting of the Cognitive Science Society, Boston, MA

Green MJ, van Deemter K (2013) The utility of vagueness: Does it lie elsewhere? In: Production of referring expressions: Bridging the gap between cognitive and computational approaches to reference. 31 July 2013 Berlin, Germany

Hobbs JR (1985) Granularity. In: Proceedings of the Ninth International Joint Conference on Artificial Intelligence, Morgan Kaufmann, pp 432–435

Izard V, Dehaene S (2008) Calibrating the mental number line. Cognition 106(3):1221–1247

Jaeger T (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language 59(4):434–446

Keefe R, Smith P (eds) (1997) Vagueness: a Reader. A Bradford Book. The MIT Press, Cambridge, MA

Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in Psychtoolbox-3. Perception 36

Kuznetsova A, Bruun Brockhoff P, Haubo Bojesen Christensen R (2016) lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). URL http://CRAN.R-project.org/package=lmerTest, version 2.0-32

Levy R (2014) Using R formulae to test for main effects in the presence of higher-order interactions. ArXiv e-prints 1405.2094

Lipman BL (2000) Comments section. In: Rubinstein A (ed) Economics and language: Five essays, Cambridge University Press

Lipman BL (2009) Why is Language Vague?, retrieved 12 April 2011 from http://people.bu.edu/ blipman/Papers/vague5.pdf

Mishra H, Mishra A, Shiv B (2011) In praise of vagueness: Malleability of vague information as a performance-booster. Psychological Science 22(6):733–738

Peters E, Dieckmann N, Västfjäll D, Mertz C, Slovic P, Hibbard J (2009) Bringing meaning to numbers: The impact of evaluative categories on decisions. Journal of Experimental Psychology: Applied 15(3):213

Trick L, Pylyshyn Z (1994) Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. Psychological Review 101:80–102

van Deemter K (2009) Utility and Language Generation: The Case of Vagueness. Journal of Philosophical Logic 38(6):607–632

van Deemter K (2010) Not Exactly: in Praise of Vagueness. Oxford University Press, Oxford, UK

van Deemter K (2016) Computational Models of Referring: a Study in Cognitive Science. MIT Press, Cambridge, Mass

van Oeffelen M, Vos P (1982) A probabilistic model for the discrimination of visual number. Perception and Psychophysics 32(2):163–170

van Rooij R (2003) Being polite is a handicap: Towards a game theoretic analysis of polite linguistic behavior. In: Tenneholz M (ed) TARK 9: Theoretical Aspects of Rationality and Knowledge, Bloomington

Zadeh LA (1965) Fuzzy sets. Information and Control 8(3):338–353