*Journal of Behavioral Decision Making*

WILEY

RESEARCH ARTICLE OPEN ACCESS

# When Half Is at Least 50%: Effect of "Framing" and Probability Level on Frequency Estimates

David R. Mandel | Megan Kelly

Intelligence, Influence, and Collaboration Section, Toronto Research Centre, Defence Research and Development Canada, Toronto, Ontario, Canada

**Correspondence:** David R. Mandel (david.mandel@forces.gc.ca; drmandel66@gmail.com)

**Received:** 22 December 2022 | **Revised:** 12 June 2024 | **Accepted:** 12 June 2024

**Funding:** This research was funded by the Canadian Safety and Security Program project CSSP-2018-TI-2394 and the Department of National Defence Accelerate Command, Control, and Intelligence Project Activity AC2I-019.

**Keywords:** at least | framing | frequency estimation | linguistic quantifiers | probability

## ABSTRACT

Expert judgment often involves estimating magnitudes, such as the frequency of deaths due to a pandemic. Three experiments (*N*s = 902, 431, and 755, respectively) were conducted to examine the effect of outcome framing (e.g., *half* of a threatened group expected to survive vs. die), probability level (low vs. high), and probability format (verbal, numeric, or combined) on the estimated frequency of survivals/deaths. Each experiment found an interactive effect of frame and probability level, which supported the hypothesis that forecasted outcomes received by participants were implicitly quantified as lower bounds (i.e., "*at least* half"). Responding in a manner consistent with a lower-bound "at least" interpretation was unrelated to incoherence (Experiments 1 and 2) and positively related to numeracy (Experiments 1 and 3), verbal reasoning (Experiment 3), and actively open-minded thinking (Experiments 2 and 3). The correlational results indicate that implicit lower bounding is an aspect of linguistic inference and not a cognitive error. Implications for research on framing effects are discussed.

## 1 | Introduction

Experts are often called upon to provide assessments of an expected frequency or magnitude. For instance, marine biologists might be called on to estimate the effect of a 1° rise in sea temperature on the number of particular fish species in a given region. Or, intelligence analysts might be asked to estimate the number of civilians who might die if a military intervention under consideration were to be conducted. In such cases, experts can seldom estimate or forecast quantities with certainty. Often, they will provide probability estimates of expected frequencies, which may be expressed in proportional terms. The marine biologist might provide an assessment such as, "if sea temperature were to rise by one degree in the next year, it is highly likely that the current populations of several fish species will be cut to *half* of their current levels within the next 5 years."

Such assessments have multiple uncertainties associated with them. Perhaps most transparently, there is the first-order probability, expressed in this example with the term *highly likely*. However, verbal probabilities are themselves vague (for a review, see Dhami and Mandel 2022), creating "decoding uncertainties" for communication receivers (Ho et al. 2015). The assessment also involves quantification that receivers must interpret. For instance, is the term *half* in the marine biologist example to be interpreted as *exactly* half, *at least* half, *at most* half, or yet some other way? Linguistic accounts of quantifiers vary in their predictions (for a review, see Spector 2013), yet all permit circumstances in which each of the former *approximators* (Ferson et al. 2015) might be expected. However, in most accounts, the *at most* interpretation is neither part of the default semantics nor standard pragmatic rules but rather licensed by knowledge of circumstances within which the expression arises. According to neo-Gricean accounts

(Horn 1992; Levinson 2000; van Rooij 2006), quantifiers are lower bounded (i.e., their default semantics is "at least *q*," where *q* stands for quantifier), but this reading could be pragmatically strengthened to convey "exactly *q*" (i.e., to convey "at least *q and* not more than *q*"). In contrast, Breheny (2008) posits an "exactly *q*" semantics that can be pragmatically weakened to "at least *q*." Geurts (2006) posits that quantifiers are ambiguous and can take on "at least *q*" or "exactly *q*" semantics, whereas Carston (1998) argues that quantifiers are not only ambiguous but also underspecified. More generally, the alternative accounts differ in the importance that they assign to compositional (grammatical) or postcompositional (pragmatic) linguistic processes (Chierchia, Fox, and Spector 2012).

Remarkably, literature in judgment and decision-making, which frequently uses tasks involving descriptions with quantifiers, has all but neglected the important implications of linguistic theories of quantification for understanding framing effects (Mandel 2015b) and other results that bear on claims about human rationality in judgment and choice (Borg 2022). This is remarkable precisely because a framing effect *is* a linguistic effect in which alternative descriptions conveying the same information (deemed "descriptively invariant") due to their extensional equivalence still have different effects on preference, choice, or judgment (Tversky and Kahneman 1981). Nevertheless, researchers for the most part have simply assumed that quantifiers are interpreted as "exactly the quantity specified."

Our principal aim here is not to test alternative linguistic theories of quantification but, rather, to leverage their permissibility of lower bounding (i.e., *at least* readings) as an explanatory construct. This construct is then used to test predictions about how the framing of the nonnumeric quantifier, *half*, and the expression of probabilities assigned to quantified outcomes (as in the marine biologist example given earlier) will interact to influence the interpretation of such statements. We build directly on framing research which has shown that one reason why gain–loss framing effects occur is due to the lower bounding of quantifiers in alternatively framed prospects. Mandel (2014, Experiment 3) demonstrated that the majority of participants, when given a standardly worded version of an isomorph of the well-known Asian Disease Problem (Tversky and Kahneman 1981), interpreted the numeric quantifiers in the certain option to mean "*at least* the specified quantity." In the same experiment, among participants who adopted an *exactly* interpretation of the relevant quantifiers, no framing effect was observed. However, a framing effect was found among participants having a lower-bound (*at least*) interpretation. These findings suggest that past theoretical accounts of framing effects are incomplete and call into question prescriptive claims about the (ir)rationality of human decision-making (Fisher and Mandel 2021; Mandel 2022; Teigen 2011).

## 2 | The Present Research

This work extends the investigation of the role of lower bounding in framing effects. In contrast to Mandel (2014), the present research focuses on the framing of outcomes under conditions in which the quantifier is nonnumeric. Specifically, we examined the interpretation of the quantifier *half* in conjunction with probability information that was presented in alternative formats including numeric, verbal, or a combination of the two. Whereas earlier work focused on the effect of frames on risky choice, this research focused on the effect of frames on frequency estimation. To illustrate the issue more clearly, consider a scenario in which 1000 lives are threatened by war. We manipulated frames by asking participants to consider assessments worded either as "It is likely that half of these civilians will survive" or as "It is likely that half of these civilians will die," after which they were asked to estimate the exact number of people who will be saved or who will die, respectively. (For purposes of analysis, we converted estimates of the number of deaths in the die condition into complementary estimates of lives saved by subtracting the original value provided from 1000.) If the quantifier *half* is interpreted as *exactly* half, then in both frames, a participant asked to estimate the frequency of survivors should indicate 500. If individuals adopt an *at least* lower-bound interpretation of the quantifier, reading it as "certainly half and possibly more" (Geurts and Nouwen 2007), then, on average, the estimate of survivors should be *greater* than 500 in the *survive* condition and *less* than 500 in the *die* condition.

We further expected that the effect of lower-bounded interpretations of alternative frames would interact with the stated probability level. In contrast to the high probability condition just considered, imagine that the assessment given had instead used the term *unlikely*, namely, "It is unlikely that half of these civilians will [survive, die]. If *half* were treated as "exactly 500," the statement might convey a symmetric displacement from the quantity. That is, the improbability of "exactly 500" surviving or dying could result in either more or less than 500 surviving or dying. If there were no preference for one or the other, one might observe a bimodal distribution with a mean estimate close to 500. However, if participants tend to interpret the quantifier as "at least 500"—namely, as 500 or *more*—and if this possible range is *unlikely*, one might expect the mean estimate of survivors to be *less* than 500. Following the same reasoning, we expected that when it is *unlikely that half will die*, the mean estimated number of survivors would exceed 500 (i.e., less than half are expected to die). We tested the predicted interaction effect between frame and probability level in three experiments.

Our research also examined whether this predicted interaction effect is moderated by probability format. For instance, some studies suggest that verbal probabilities signal optimism or pessimism about future states more effectively than numeric probabilities of comparable magnitude (Collins and Mandel 2019; Collins, Mandel, and Macleod 2024; Teigen and Brun 1995). Accordingly, one might predict that the hypothesized interaction effect just described would be stronger in a verbal probability format than in a numerical format. However, the interpretation of verbal probabilities is often regressive (e.g., Mellers et al. 2017; Mandel 2015a; Mandel and Irwin 2021; Smithson et al. 2012) and fuzzy (for a review, see Dhami and Mandel 2022). Regression could result in a weaker interaction effect in the verbal condition because it would weaken the manipulation of the probability level relative to the numeric condition.

Finally, in contrast to the linguistic accounts of quantification noted earlier, some decision-making literature suggests that participants who do *not* adopt the *exactly* interpretation of quantifiers may have adopted an inferior interpretation. For instance, Shafir and LeBoeuf (2002) state that some coherence violations might be due to "conversational misinterpretations" (p. 506). This "deficient reading" hypothesis is not well supported by existing evidence. Susceptibility to framing effects is not robustly correlated with measures of cognitive ability or cognitive styles such as need for cognition and actively open-minded thinking (AOT) (Mandel and Kapler 2018). Nevertheless, no study, to our knowledge, has examined whether so-called deficient readings of quantifiers, such as interpreting the term *half* as a lower bound, are associated with indicators of irrational judgment or choice. We tested this hypothesis by examining whether various task-related coherence violations (e.g., violations of monotonicity) were greater among participants whose magnitude judgments conformed to a lower-bound interpretation than to those whose judgments did not conform to such an interpretation. We expected that coherence violations would be inversely related to other indicators of cognitive ability, such as numeracy, verbal reasoning skill, and an AOT disposition (Baron et al. 2015), which has been negatively associated with accuracy (Haran, Ritov, and Mellers 2013) and certain cognitive biases that violate coherence principles (Toplak, West, and Stanovich 2017). However, in line with linguistic account of quantification, we expected that such violations would be uncorrelated with whether participants gave lower-bound-consistent estimates or not.

## 3 | Experiment 1

### 3.1 | Materials and Methods

#### 3.1.1 | Participants

Participants were recruited from Canada and the United States through Qualtrics Panels, an online crowdsourcing service, at an estimated remuneration rate of 7.50 USD. Participants were required to have English as their first language and were prohibited from completing the experiment on a smartphone. Consistent with prior research (Mandel and Irwin 2021; Irwin and Mandel 2023), participants were screened out if they failed to pass a one-item instructional manipulation check used to detect inattention to task instructions (Oppenheimer, Meyvis, and Davidenko 2009) or if they provided incoherent lower- and upper-bound estimates during a probability-interpretation task (i.e., if their lower-bound estimate exceeded their relevant upper-bound estimate). The sample ($N = 902$) meeting these criteria was 50.8% male and had a mean age of 43.82 years ($SD = 11.66$). This sample size exceeded the calculation of a minimum required sample size of 760 for the primary analysis involving factorial analysis of variance (ANOVA) with 12 between-subject conditions (i.e., probability format [3] × probability level [2] × frame [2]). Using G*Power (Faul et al. 2007), this estimate was based on setting Type I and II error rates set to 5% and using an estimated partial $\eta^2 = 0.02$ (i.e., a small- to medium-sized effect for the predicted two-way interaction). The larger sample obtained was sufficient to meet the requirements of other statistical tests unrelated to this research, and the postscreened sample size was set a priori with the Qualtrics Panels service provider.

#### 3.1.2 | Design

Participants were randomly assigned to 12 conditions in a 3 (probability format: verbal, numeric, combined) × 2 (probability level: low, high) × 2 (frame: survive, die) between-subject factorial design. Probability format refers to whether participants responded to an intelligence forecast containing a verbal probability term (e.g., *likely*), a numeric range (e.g., 60%–90%), or a verbal term combined with a parenthesized numeric range (e.g., *likely* [60%–90%]). Probability level indicates whether the forecast conveyed a low probability (e.g., *unlikely* [10%–40%]) or a high probability (e.g., *likely* [60%–90%]). Frame refers to whether the forecasted outcome was framed positively (i.e., half of a group of displaced civilians will survive) or negatively (i.e., half of the group will die).

#### 3.1.3 | Materials and Procedure

Supporting Information and data for Experiments 1–3 are available at https://osf.io/m8kst/.

Prior to commencement, Defence Research and Development Canada's Human Research Ethics Committee approved Experiments 1–3. Experiment 1 was administered using Qualtrics and counterbalanced with two other brief studies (described in Collins and Mandel 2019 and Mandel et al. 2021, respectively). During the experiment, participants were unable to view or alter responses entered on previous screens.

At the start of Experiment 1, participants were informed that they would answer a series of questions regarding a hypothetical intelligence forecast. Parts of the procedure were intended to address questions about the effect of probability format on the interpretation of probability assessments. Mandel and Irwin (2021) discuss those research aims. Participants were introduced to the NATO standard for communicating probabilities in intelligence assessments, and they were told that an analyst had used one of the probability terms when making the forecast. The use of the standard was not central to the present research aims. However, we report the method in full. A vignette containing a hypothetical intelligence forecast was then presented as follows (probability format indicated in braces, probability level, and frame manipulations shown in brackets):

In a war-torn region, the UN recently relocated 1000 displaced civilians from a particular ethnic group to a safer location. However, their lives are now at stake due to the expansion of ethnopolitical warfare into that region.

Given the current situation on the ground, a senior intelligence analyst specializing in that region assesses

> {Verbal condition} "It is [likely/unlikely] that half of these civilians will [survive/die]."

> {Numeric condition} "There is a [60% - 90%/10% - 40%] chance that half of these civilians will [survive/die]."

> {Combined condition} "It is [likely (namely, there is a 60% - 90% chance)/unlikely (namely, there is a 10% - 40% chance)] that half of these civilians will [survive/die]."

The text above remained onscreen as participants answered questions about the forecast. Participants were able to review the NATO standard by clicking a button before proceeding to the first set of questions. During the subsequent probability-interpretation task, participants were asked to provide their best, lowest, and highest estimates of the probability that the intelligence analyst had in mind using sliders ranging from 0 to 100. (All sliders in the core experiment had a default starting position of zero. Participants who wanted to select zero as their response were still required to input it by clicking the slider.) Mandel and Irwin (2021) analyzed responses to these questions, which were not the focus of the present research. However, as noted earlier, these data were used to exclude participants who provided incoherent lower- and upper-bound estimates (i.e., where the former was larger than the latter).

On the same page, participants were asked a fourth question (frame manipulation shown in brackets): "If you had to estimate the **exact number** of civilians that will [survive/die], what would it be?" They responded on a slider ranging from 0 to 1000, which moved in increments of 1 and had labels for each increment of 100. This response variable, after transformation (described later), served as the primary dependent variable.

Participants were once again able to review the NATO standard. The next series of questions probed participants' judgments of cumulative probability. Using sliders ranging from 0 to 100, they were asked to estimate the probability that 0 or more civilians will [survive/die], 100 or more civilians will [survive/die], 200 or more civilians will [survive/die], and so on (i.e., 300, 400, 500, 600, 700, 800, 900, or more) up to the probability that all 1000 civilians will [survive/die]. Data from this task were used to compute two coherence measures. First, whether participants assigned a probability of 100% to the possibility of "0 or more" surviving/dying was recorded. Because "0 or more" includes all possibilities, a coherent response would be to assign a 100% chance. Second, the number of monotonicity violations committed by each participant within his or her set of cumulative probability judgments was recorded. The two measures had acceptable scale reliability (Cronbach's $\alpha = 0.67$) and were averaged following standardization. Higher values on the scale indicate greater incoherence.

Next, participants completed the one-item instructional manipulation check (Oppenheimer, Meyvis, and Davidenko 2009) which, as noted earlier, was used as a screening criterion. Those who successfully completed this item then completed in the following order: (a) a 10-item numeracy scale comprising eight questions from Lipkus, Samsa, and Rimer (2001) and two questions from Cokely et al. (2012); (b) an 8-item verbal reasoning skill test using verbal analogy items from the 29-item Penn Verbal Reasoning Test (Bilker et al. 2014); and (c) the eight-item AOT scale from Baron et al. (2015). Participants responded to the AOT items on a 5-point scale ranging from −2 (*strongly disagree*) to 2 (*strongly agree*). The numeracy and verbal reasoning tests were scored as the proportion of correct responses, and AOT was calculated as the mean across items. Finally, participants provided demographic data used to describe the sample.

### 3.2 | Results

Frequency estimates in the die condition were converted to the estimated number of *survivors* by subtracting participants' estimates from 1000. We then conducted a 3 (probability format)×2 (probability level)×2 (frame) between-subject factorial ANOVA on the estimated number of lives saved using $\alpha = 0.05$ as the Type I error rate. Levene's test of the equality of error variances was significant ($F[11, 890] = 3.12$, $p < 0.001$). While the results must be interpreted cautiously, given that the design-cell sample sizes did not vary substantially ($n_{max}/n_{min} = 86/66 = 1.3$), the homogeneity of error variance assumption may be relaxed. (A heuristic for making this determination is that $n_{max}/n_{min} < 1.5$.) As expected, there was a significant frame × probability level interaction effect ($F[1, 890] = 157.25$, $p < 0.001$, $\eta_p^2 = 0.150$), which was qualified by the three-way interaction effect ($F[2, 890] = 4.54$, $p < 0.011$, $\eta_p^2 = 0.010$). (There was a significant main effect of frame ($F[1, 890] = 25.63$, $p < 0.001$, $\eta_p^2 = 0.028$). However, this effect is not of theoretical interest, and furthermore, it was qualified by the reported interactions).

Characteristics of the two-way interaction strongly cohere with a lower-bound interpretation of *half*. Consistent with lower bounding, when probabilities were described as high, the mean estimated number of survivors in the survive condition was greater than exactly half ($M = 553.90$, 95% CI [530.86, 576.93]) and the mean estimate was lower than exactly half in the die condition ($M = 465.21$, 95% CI [441.84, 488.57]). Moreover, in line with the lower-bounding hypothesis, for low probabilities, the opposite pattern occurred: the mean estimate was lower than exactly half in the survive condition ($M = 424.45$, 95% CI [400.95, 447.94]), and it was higher than exactly half in the die condition ($M = 633.24$, 95% CI [610.02, 656.46]). The three-way interaction plotted in Figure 1 shows that the crossover pattern expected on the basis of lower bounding is stronger when numeric probabilities are presented (i.e., combined and numeric conditions) than when only verbal probabilities are presented. However, without exception, all 12 cell means were in the expected direction.

Next, the lower-bounding hypothesis was tested by coding whether each participant's untransformed estimate was consistent (≥ 500) or inconsistent (< 500) with a lower-bound interpretation. Roughly two-thirds of the sample (68.7%, 95% CI [66.0%, 71.5%]) gave lower-bound-consistent estimates. This rate was not contingent on frame, probability level, or probability format (all $p > 0.10$ based on chi-square tests). Results were comparable if the minority (14.9%) of participants who gave estimates of exactly 500 were excluded. Among the remaining 768 participants, 68.8% (95% CI [65.5%, 71.6%]) gave lower-bound consistent responses.

Finally, the normative basis of lower-bound responding was probed by examining whether participants who conformed to this pattern were more prone to committing coherence violations

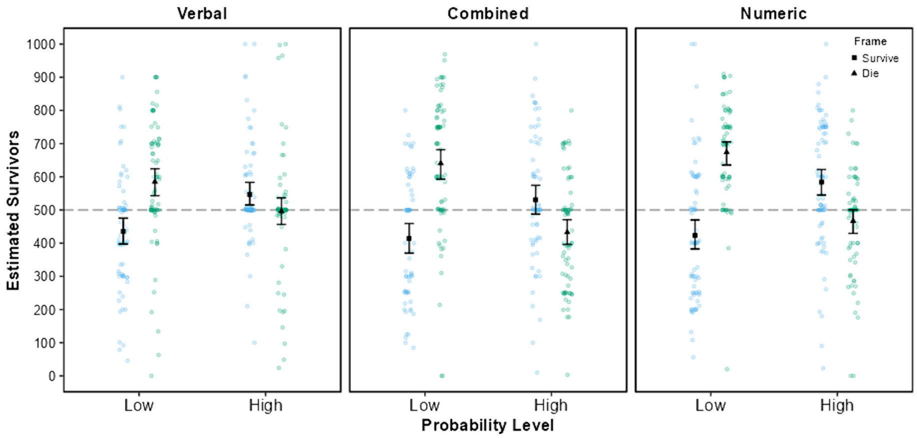


**FIGURE 1** | Estimated frequency of survivors by probability format, probability level, and frame in Experiment 1. *Note:* Error bars are estimated marginal means with 95% confidence intervals. Raincloud plots are from sample data.

**TABLE 1** | Pearson correlations among key variables in Experiment 1.

| | LBC | NUM | VR | AOT | INCOH |
|---|---|---|---|---|---|
| LBC | — | 0.13** | 0.04 | −0.01 | −0.04 |
| NUM | **0.13**** | 0.69 (0.21) | 0.45** | 0.36** | −0.34** |
| VR | **0.03** | | 0.56 (0.24) | 0.37** | −0.26** |
| AOT | **−0.03** | | | 0.67 (0.57) | −0.25** |
| INCOH | **−0.05** | | | | 0.00 (0.87) |

*Note:* Scale *M* and *SD* (the latter are in parentheses) are reported on the diagonal line. The scores for NUM and VR scales are the proportion of items correct. Bold values are correlations that remove participants who estimated a value of 500 (remaining $n=768$).
Abbreviations: AOT = actively open-minded thinking scale, INCOH = incoherence scale, LBC = lower-bound consistent (0 = *no*, 1 = *yes*), NUM = numeracy scale, VR = verbal reasoning scale.
$^*p<0.05$, and $^{**}p<0.01$.

within the task and whether lower-bound responders were less numerate, less adept at verbal reasoning, or less prone to having an AOT style. A majority (75.8%) gave an incoherent response to the question about the probability of zero or more surviving and a majority (60.5%) exhibited at least one monotonicity violation. As Table 1 shows, the tendency to provide estimates consistent with lower bounding was weakly *positively* related to numeracy, but unrelated to any of the other measures. Participants who gave lower-bound-consistent estimates tended to be more numerate. The findings were comparable if participants who responded with a value of 500 were excluded (see bold values in Table 1). As expected, incoherence was associated with lower numeracy, poorer verbal reasoning, and a lower propensity for AOT.

### 3.3 | Discussion

Both the analysis of mean differences in frequency estimates and proportions of lower-bound-consistent responding in Experiment 1 supported the hypothesis that many individuals treat quantifiers in descriptions as lower bounds, or less restrictively, they do not necessarily treat quantifiers as meaning "exactly that much." Lower-bound-consistent responding was associated with greater numeracy and not with greater incoherence, contrary to the hypothesis that interpretations of quantifiers that do not align with an "exactly" reading represent a form of cognitive deficiency.

## 4 | Experiment 2

Experiment 2 aimed to replicate the key findings of Experiment 1 while implementing methodological improvements. In Experiment 1, participants were asked to interpret the probability expression contained in the intelligence forecast before making their frequency estimate. The former task response could have influenced their estimation responses, which are critical to the present hypothesis tests. In Experiment 2, task order was modified such that the measures of primary interest (frequency estimates and the cumulative probability judgments) were completed first. Experiment 2 was also simplified by focusing on the effect of two distinct probability formats (verbal and numeric) and by removing opportunities for participants to review the NATO standard on the communication of probabilities, which had been implemented in Experiment 1 for purposes unrelated to the present research aims.

### 4.1 | Materials and Methods

#### 4.1.1 | Participants

Native English speakers were recruited from Canada and the United States through Qualtrics Panels using the same screening criteria applied in Experiment 1. The sample ($N=431$) meeting these criteria was 50.6% male with a mean age of 41.71 years ($SD=11.57$). The sample slightly exceeded the estimate of 423 required to ensure comparable statistical power to Experiment 1 with a slightly increased estimated effect size (partial $\eta^2=0.03$), which may be regarded as a conservative estimate given the larger one obtained in Experiment 1 for the predicted two-way (frame × probability level) interaction effect. As in Experiment 1, the required postscreened sample size was communicated to the Qualtrics Panels service provider in advance of data collection.

#### 4.1.2 | Design

Participants were randomly assigned to eight conditions in a 2 (probability format: verbal, numeric) × 2 (probability level: low, high) × 2 (frame: survive, die) between-subject factorial design.

#### 4.1.3 | Materials and Procedure

Experiment 2 was counterbalanced with another brief task described in Collins, Mandel, and Macleod (2024). As in Experiment 1, participants were first introduced to the NATO standard for communicating probability in intelligence assessments. Next, they were presented with the same vignette used in Experiment 1. Participants then estimated the exact number of civilians that would survive or die depending on the framing condition to which they were assigned. The question wording and response scale were identical to those used in Experiment 1. Participants then completed the cumulative probability judgment task, followed by the probability-interpretation task. (After providing probability interpretations, participants were asked "How confident are you that the analyst's intended probability falls within your lower and upper bounds?" They responded on a slider ranging from 50 ("Not at all confident") to 100 ("Completely confident"), which moved in increments of 1. Data from the probability interpretation task (including the confidence question) were collected for an unrelated set of aims closely related to those described by Mandel and Irwin (2021).

Following the core experimental tasks, participants who passed the instructional manipulation check went on to complete the numeracy scale (Whereas the numeracy test used in Experiment 1 elicited a combination of multiple-choice and text-box inputs, in Experiment 2, we used a purely multiple-choice test for ease of scoring. For both versions of the numeracy test, see https://osf.io/m8kst/), the verbal analogy items from the Penn Verbal Reasoning Test, the AOT scale, and demographic questions in that order.

### 4.2 | Results

As in Experiment 1, frequency estimates in the die condition were converted to the estimated number of survivors by subtracting them from 1000. A 2 (probability format) × 2 (probability level) × 2 (frame) between-subject factorial ANOVA was conducted on the estimated number of survivors using $\alpha = 0.05$ as the Type I error rate. Levene's test was significant ($F[7, 423] = 2.34$, $p = 0.024$). However, as in Experiment 1, the sample sizes of the specific conditions did not vary substantially ($n_{max}/n_{min} = 65/49 = 1.3$) and the requirement for the homogeneity of variance assumption to be met may be relaxed. There was a significant frame × probability level interaction effect ($F[1, 423] = 63.33$, $p < 0.001$, $\eta_p^2 = 0.130$), which was qualified by three-way interaction ($F[1, 423] = 4.68$, $p = 0.031$, $\eta_p^2 = 0.011$). (There were also significant main effects of presentation format ($F[1, 423] = 6.66$, $p = 0.010$, $\eta_p^2 = 0.016$) and frame ($F[1, 423] = 7.54$,

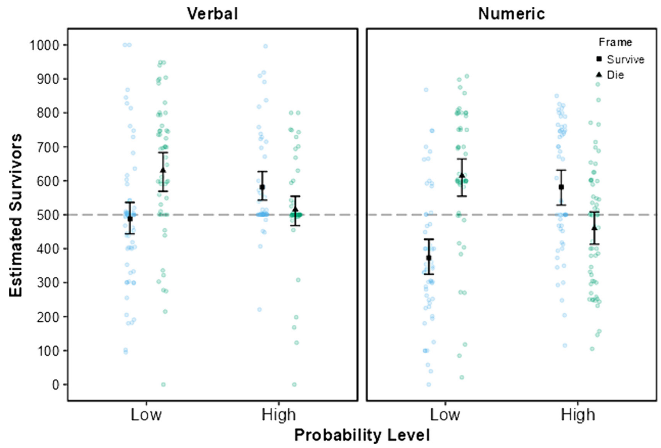


**FIGURE 2** | Estimated frequency of survivors by probability format, probability level, and frame in Experiment 2. *Note:* Error bars are estimated marginal means with 95% confidence intervals. Raincloud plots are from sample data.

$p = 0.006$, $\eta_p^2 = 0.018$). However, these main effects are not of theoretical interest and were furthermore qualified by the three-way interaction effect.) Supporting the lower-bound hypothesis, when the probability level was high, the mean estimate in the survive condition was greater than exactly half ($M = 581.30$, 95% CI [545.14, 617.46]) and the mean estimate was less than exactly half in the die condition ($M = 487.72$, 95% CI [452.36, 523.08]), although in the latter case, the confidence interval includes the exactly half value of 500. When the probability level was low, the mean estimate in the survive condition was lower than exactly half ($M = 430.54$, 95% CI [397.11, 463.97]) and the mean estimate was greater than exactly half in the die condition ($M = 622.72$, 95% CI [586.58, 658.85]). As can be seen in Figure 2, the results are consistent with Experiment 1 in showing that the pattern of results expected based on the lower-bounding hypothesis is clearer in the numeric condition. In the numeric condition, three of the four cells have means that significantly differ from exactly half in the expected direction (the mean in the high-die condition that did not significantly differ was nevertheless in the prediction direction). In contrast, in the verbal condition, both of the cells with means expected to be less than exactly half included the reference point in the 95% confidence intervals, and in one case (once again, the high-die condition), the mean estimate was in the opposite direction to that expected based on the hypothesis.

As in Experiment 1, the lower-bounding hypothesis was also tested by coding whether each participant's estimate was consistent or inconsistent with a lower-bound interpretation. Just over two-thirds of the sample (68.9%, 95% CI [64.7%, 73.1%]) gave lower-bound-consistent estimates. This rate was not contingent on frame, probability level, or probability format (all $p > 0.80$ based on chi-square tests). If participants whose estimates equalled exactly 500 ($n = 43$) were excluded, the majority providing lower-bound-consistent estimates was comparable ($n = 388$, 67.3%, 95% CI [62.6%, 71.9%]).

Finally, we examined whether participants who conformed to a lower-bound pattern of estimation were more prone to committing coherence violations within the task and whether

**TABLE 2** | Pearson correlations among key variables in Experiment 2.

| | LBC | NUM | VR | AOT | INCOH |
|---|---|---|---|---|---|
| LBC | — | 0.08 | 0.06 | 0.10* | −0.04 |
| NUM | **0.07** | 0.71 (0.21) | 0.43 | 0.39** | −0.36** |
| VR | **0.02** | | 0.50 (0.24) | 0.43** | −0.36** |
| AOT | **0.07** | | | 0.60 (0.59) | −0.28** |
| INCOH | **−0.03** | | | | 0.00 (0.88) |

*Note:* Scale *M* and *SD* (the latter are in parentheses) are reported on the diagonal line. The scores for NUM and VR scales are the proportion of items correct. Bold values are correlations that remove participants who estimated a value of 500 (remaining $n = 388$).
Abbreviations: AOT = actively open-minded thinking scale, INCOH = incoherence scale, LBC = lower-bound consistent (0 = *no*, 1 = *yes*), NUM = numeracy scale, VR = verbal reasoning scale.
*$p < 0.05$, and **$p < 0.01$.

lower-bounders were less numerate, adept at verbal reasoning, or less prone to having an AOT style, as one might expect on the basis of the deficient-reading hypothesis. The method was identical to that used in Experiment 1 and Cronbach's $\alpha = 0.70$ for the coherence measure. A majority (74.2%) gave an incoherent response to the question about the probability of zero or more surviving, and a majority (65.7%) exhibited at least one monotonicity violation. As Table 2 shows, lower-bound-consistent responding was weakly *positively* related to AOT, but unrelated to any of the other measures. As in Experiment 1, the findings were comparable if participants who responded with a value of 500 were excluded (see bolded values in Table 2). As in Experiment 1, task-related incoherence was associated with lower numeracy, poorer verbal reasoning skill, and lower AOT.

### 4.3 | Discussion

Experiment 2 replicated the key results of Experiment 1 both in terms of the analysis of mean differences in frequency estimates and the proportion of lower-bound-consistent responses. Moreover, such responses were not associated with incoherence and showed a weak positive relation with an AOT style. Taken together, these results supported the lower-bounding hypothesis and did not support the cognitive deficiency hypothesis.

## 5 | Experiment 3

Experiment 3 was built on the previous experiments in several respects. First, we assumed in the prior experiments that the target population of 1000 was interpreted as an exact amount, yet the premise of this research is that a quantifier *n* is not necessarily interpreted as *exactly n*. In Experiment 3, therefore, we made it explicit that the reference population of 1000 referred to "exactly 1000." Second, we used a wider range of scenarios to test the robustness of the observed effects. One of the three scenarios focused on the number of male or female births recorded. Because the expected proportions of male and female births are equiprobable and this is a widely known fact, we reasoned that there might be a weaker tendency to lower bound the quantifier *half* in this scenario. Third, we omitted the probability translation task used in Experiments 1 and 2, which, as noted earlier, had been implemented for a separate report. This obviated the need for presenting other unrelated material such as the NATO standard for communicating verbal probabilities, which was shown to participants in Experiments 1 and 2. We also used different probability terms and corresponding values to further increase the generalizability of the findings. Finally, we queried participants about their linguistic interpretation of "half" by asking whether they interpreted the quantifier as meaning either *at least*, *at most*, *exactly*, or *roughly* that much. Although linguistic accounts of quantifiers often focus on single-bounded or exact readings of such terms (Spector 2013), everyday experience tells us that they are also often interpreted as rough approximations such as "about half" (Ferson et al. 2015), especially when the quantifiers are round such as "half of 1000" (Sadock 1977).

### 5.1 | Materials and Methods

#### 5.1.1 | Participants

As in the previous experiments, native English speakers were recruited from Canada and the United States using Qualtrics Panels with the same screening criteria applied. The sample ($N = 755$) meeting these criteria was 48.7% male with a mean age of 43.5 years ($SD = 11.20$). (One participant (female) did not provide a response to the age demographic question and was, therefore, not included in the computations of mean and *SD* of age.) The sample size substantially exceeded the estimate of 384 required to ensure comparable statistical power to earlier experiments with an estimated effect size of partial $\eta^2 = 0.03$. As in the prior experiments, the desired postscreened sample size was communicated to the Qualtrics Panels service provider prior to data collection.

#### 5.1.2 | Design

Participants were randomly assigned to eight conditions in a 2 (probability format: verbal, numeric) × 2 (probability level: low, high) × 2 (frame: A, B) between-subject factorial design. The verbal terms used in the low and high probability conditions were *very low probability* and *very high probability*, respectively. In the numeric condition, the corresponding values were "less than a 10% probability" and "greater than a 90% probability," respectively. Experiment 3 also included a within-subject factor of scenario wherein participants each completed the tasks in response to three different scenarios. Specifically, there was a scenario that used the *survive/die* dichotomy and which was similar to the scenario used in the previous experiments (refugee scenario), a scenario that used a *remain/removal* dichotomy with respect to the number of species on an endangered species list (species scenario), and a scenario regarding the number of *male/female* births at a hospital (birth scenario). Scenarios are presented in full in the supplementary materials. If participants were in the frame A condition, then they received the *survive* framing for the refugee scenario, the *remain* framing for the species scenario, and the *male* framing for the birth scenario. In contrast, participants in the frame B condition received the *die* framing (refugees), the *removal* framing (species), and the *female* framing (births).

### 5.1.3 | Materials and Procedure

For each scenario, participants completed a frequency estimation task and a quantifier interpretation task, the order of which was randomized across participants. The order of the scenarios was also randomized for each participant. The estimation task was comparable in wording to that used in the previous experiments. For example, in the refugee scenario as an example, the frequency estimation question was phrased, "If you had to estimate the precise number of refugees that will survive through the end of the winter season, what would it be?" As in the previous experiments, participants responded on a 1001-point slider scale ranging from 0 to 1000. For the quantifier interpretation task, participants were asked (in the refugee scenario), "In the assessment, how do you think the term *half* as in 'half of these refugees' is best understood?"They responded by selecting either "exactly half," "at most half," "at least half," or "roughly half." The order of these options was randomized per participant. Participants responded to both questions for a scenario before moving on to the next scenario (see supplementary materials). After completing the core experimental tasks, participants were given the instructional manipulation task as in Experiment 2. Those who passed the screening test also completed the numeracy scale, the verbal analogy items from the Penn Verbal Reasoning Test, and the AOT scale and answered demographic questions as in Experiment 2.

## 5.2 | Results

Frequency estimates in frame B conditions were converted to the estimated number of survivors, male births, and species remaining on the endangered species list (i.e., frame A) by subtracting estimates from 1000. A 2 (probability format)×2 (probability level)×2 (frame)×3 (scenario) mixed ANOVA was conducted on the estimated frequencies using $\alpha = 0.05$ as the Type I error rate. Levene's test was significant for each of the three scenarios (all $p < 0.001$), yet as in the previous experiments, the sample sizes of the specific conditions did not vary substantially ($n_{max}/n_{min} = 97/88 = 1.1$). Of note, the predicted frame×probability level interaction effect was significant ($F[1, 747] = 100.06$, $p < 0.001$, $\eta_p^2 = 0.118$). (There were also significant main effects of scenario (Greenhouse–Geisser $F[2, 1463.92] = 7.27$, $p < 0.001$, $\eta_p^2 = 0.010$) and frame ($F[1, 747] = 22.16$, $p < 0.001$, $\eta_p^2 = 0.029$) and a frame×probability format interaction effect ($F[1, 747] = 3.84$, $p = 0.050$, $\eta_p^2 = 0.005$). However, these effects are not of theoretical interest and were furthermore qualified by 2 three-way interaction effects reported in the main text.) This interaction effect was qualified by 2 three-way interaction effects. First, as in Experiments 1 and 2, the frame×probability level×probability format interaction effect was significant ($F[1, 747] = 9.10$, $p = 0.003$, $\eta_p^2 = 0.012$).

Figure 3 shows that the two-way interaction conforms to the lower-bounding hypothesis and replicates the key feature of the three-way interaction from Experiments 1 and 2—namely, that the predicted interaction effect is stronger in the numeric condition than in the verbal condition. There was also a significant frame×probability level×probability format interaction effect (Greenhouse–Geisser $F[1.96, 1463.92] = 3.06$, $p = 0.048$, $\eta_p^2 = 0.004$), which is plotted in Figure S1. Although this three-way interaction is difficult to interpret (see Figure S1), two points are noteworthy. First, the predicted two-way interaction that is consistent with the lower-bounding hypothesis was statistically significant in each of the three scenarios: for the refugee scenario, $F(1, 747) = 92.27$, $p < 0.001$, $\eta_p^2 = 0.110$; for the species scenario, $F(1, 747) = 48.19$, $p < 0.001$, $\eta_p^2 = 0.061$; and for the birth scenario, $F(1, 747) = 67.73$, $p < 0.001$, $\eta_p^2 = 0.083$. Second, the effect size of this interaction was not weakest in the birth scenario, where one might have expected the quantifier to be interpreted as "exactly half" or "roughly half" rather than "at least half."

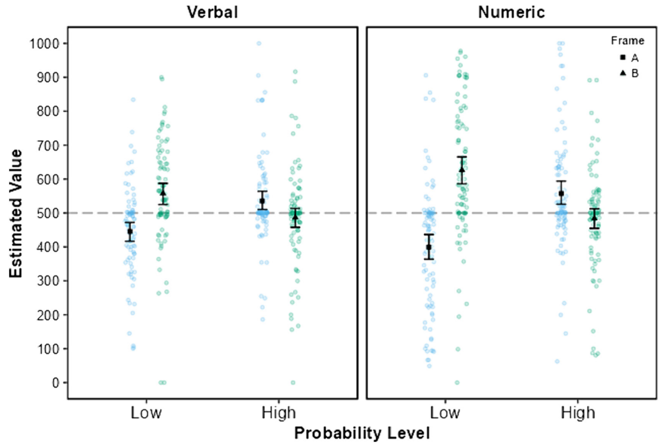


**FIGURE 3** | Estimated event frequency (coded as frame A) by probability format, probability level, and frame in Experiment 3. *Note:* Error bars are estimated marginal means with 95% confidence intervals. Raincloud plots are from sample data.

**TABLE 3** | Percentage of lower-bound consistent responses by scenario in Experiment 3.

| | Total sample ($n = 755$) | | Excluding responses of 500 | |
|---|---|---|---|---|
| **Scenario** | **%** | **95% CI** | **% (*n*)** | **95% CI** |
| Refugees | 63.0 | 59.5, 66.4 | 65.0 (686) | 61.4, 68.6 |
| Species | 60.5 | 57.4, 63.6 | 62.1 (689) | 58.4, 65.9 |
| Births | 66.9 | 63.7, 69.9 | 68.0 (687) | 64.4, 71.5 |

The lower-bounding hypothesis was also tested by coding whether each participant's estimate was consistent or inconsistent with a lower-bound interpretation. Consistent with the results of Experiments 1 and 2, a significant majority of participants provided lower-bound-consistent responses in each scenario. As Table 3 shows, this result was observed whether the "exactly half" responses of 500 were included or excluded.

Participants' responses to the question about their interpretations of "half" were more ambiguous. As Table 4 shows, across each scenario, the modal response was to indicate that *half* meant

**TABLE 4** | Percentages of quantifier interpretations by scenario and probability level in Experiment 3.

| | Overall | | Probability level | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Low | | High | | | |
| Scenario/interpretation | % | 95% CI | % | 95% CI | % | 95% CI | $\chi^{2a}$ | *p* |
| Refugees | | | | | | | 13.17 | 0.004 |
| Roughly | 39.5 | 36.2, 42.5 | 34.6 | 30.3, 38.9 | 44.2 | 39.3, 49.2 | | |
| Exactly | 14.6 | 12.3, 17.0 | 15.3 | 11.8, 19.0 | 13.9 | 10.7, 17.3 | | |
| At least | 25.6 | 22.6, 28.6 | 24.9 | 20.9, 29.2 | 26.2 | 22.0, 30.4 | | |
| At most | 20.4 | 17.6, 23.0 | 25.2 | 21.2, 29.8 | 15.7 | 12.6, 18.8 | | |
| Species | | | | | | | 21.09 | < 0.001 |
| Roughly | 40.0 | 37.0, 43.3 | 34.9 | 30.3, 39.4 | 45.0 | 40.1, 50.1 | | |
| Exactly | 15.5 | 13.1, 18.1 | 16.4 | 13.4, 19.6 | 14.7 | 11.5, 18.1 | | |
| At least | 24.8 | 22.1, 27.4 | 22.8 | 18.5, 26.8 | 26.7 | 22.8, 31.2 | | |
| At most | 19.7 | 16.8, 22.5 | 26.0 | 21.4, 30.3 | 13.6 | 10.7, 16.5 | | |
| Births | | | | | | | 33.89 | < 0.001 |
| Roughly | 42.8 | 39.3, 46.1 | 33.0 | 28.4, 37.5 | 52.4 | 47.4, 57.1 | | |
| Exactly | 19.1 | 16.3, 21.6 | 20.9 | 17.2, 24.4 | 17.3 | 14.1, 20.4 | | |
| At least | 18.7 | 16.0, 21.2 | 20.4 | 16.6, 24.4 | 17.0 | 13.5, 20.7 | | |
| At most | 19.5 | 16.8, 22.3 | 25.7 | 21.7, 29.8 | 13.4 | 10.2, 16.5 | | |

*Note:* Confidence intervals are based on 1000 bias-corrected and accelerated bootstrap samples. Chi-square tests analyze the contingency between participants' linguistic interpretations and probability level.
[a]df = 3, $N$ = 755.

"roughly half." More surprisingly, the proportions of "at least" and "at most" responses were comparable within each scenario. This might be expected to occur if many participants did not interpret the question to refer specifically to the term *half* but rather to the entire statement (i.e., that it was [e.g., unlikely, likely] that half would be the outcome). If so, the lower-bounding hypothesis would predict a contingency between linguistic response and probability level such that there would be a higher proportion of "at least" responses observed in the high probability condition than in the low probability condition since something that is *unlikely* to bring about at least *n* will bring about at most *n*. In line with this hypothesis, responses were dependent on the probability level within each scenario (see chi-square test results in Table 4). Whereas there was a slight but nonsignificant tendency for more participants to give "at most" responses than "at least" responses in the low probability condition, a stronger opposing tendency to give more "at least" responses than "at most" responses was observed in the high probability condition (see Table 4). This pattern is consistent with the lower-bounding hypothesis. (An alternative hypothesis for the higher-than-expected rate of "at most" responses suggested by one reviewer is that some participants may have confused "at most" with "more than." However, by the same token, one could argue that "at least" selections were confused with "less than." Unlike the hypothesis that we tested, we can see no way to test this in the present experiment).

We examined participants' linguistic responses in another way by focusing on the subset of participants who provided "roughly" responses and who did not give a frequency estimate of exactly 500 within a given scenario. If "roughly" is merely imprecise but symmetric (like "exactly"), we would expect no significant difference between the proportion of "roughly" responders who give frequency estimates consistent with a lower-bounding interpretation and those who give estimates consistent with an upper-bounding interpretation. Contrary to this prediction, and supporting the lower-bounding hypothesis, the proportion who gave estimates consistent with lower-bounding was a significant majority in each scenario: for the refugees scenario: $n = 270$, proportion = 0.622, 95% CI [0.561, 0.684]; for the species scenario: $n = 267$, proportion = 0.592, 95% CI [0.534, 0.648]; for the birth scenario: $n = 294$, proportion = 0.667, 95% CI [0.613, 0.726].

Finally, we examined whether the lower-bound consistent responses were correlated with individual difference measures. As can be seen in Table 5, the results were consistent with the earlier experiments. Specifically, in the refugee scenario, lower-bound consistent responding was positively correlated with AOT, and in the birth scenario, such responding was positively correlated with numeracy, verbal reasoning, and AOT. As in the earlier experiments, the findings were comparable if participants who responded with a value of 500 were excluded (see bolded values in Table 5).

### 5.3 | Discussion

The results of Experiment 3 replicated and generalized the findings of the earlier experiments. In each of the three scenarios,

**TABLE 5** | Pearson correlations among key variables in Experiment 3.

| | LBC refugees | LBC species | LBC births | NUM | VR | AOT |
|---|---|---|---|---|---|---|
| LBC refugees | — | 0.36** | 0.39** | 0.04 | 0.06 | 0.12** |
| LBC species | | — | 0.37** | 0.03 | 0.05 | 0.03 |
| LBC births | | | — | 0.09* | 0.11* | 0.09* |
| NUM | **0.07** | **0.03** | **0.11**** | 0.63 (0.22) | 0.46** | 0.36** |
| VR | **0.08*** | **0.05** | **0.12**** | | 0.44 (0.23) | 0.36** |
| AOT | **0.13**** | **0.03** | **0.09*** | | | 0.56 (0.57) |

*Note:* Scale *M* and *SD* (the latter are in parentheses) are reported on the diagonal line. The scores for NUM and VR scales are the proportion of items correct. Bold values are correlations that remove participants who estimated a value of 500 (remaining $n_{refugees} = 686$; $n_{species} = 755$; $n_{births} = 687$).
Abbreviations: AOT = actively open-minded thinking scale, LBC = lower-bound consistent (0 = *no*, 1 = *yes*), NUM = numeracy scale, VR = verbal reasoning scale.
*$p < 0.05$, and **$p < 0.01$.

mean frequency estimation was consistent with lower bounding of the quantifier *half*, and a majority of participants provided lower-bound consistent responses for each scenario, regardless of whether responses of exactly 500 were included or excluded in the analyses. While it appears that the linguistic interpretation question was interpreted differently than intended by some participants, especially by those in the low probability condition, the results are nevertheless informative. In particular, there seems to be little doubt that when given the option to select from *exactly*, *at least*, *at most*, or *roughly*, there is a dominant preference to select *roughly*. Still, among those selecting this interpretation and not providing a frequency estimate of exactly 500, a significant majority estimated values in line with lower bounding for all scenarios.

## 6 | General Discussion

Although linguistic theories posit that quantifiers can take on a range of meanings, it has been assumed in much of the experimental work in decision science that quantifiers presented in toy decision problems like the Asian Disease Problem should be interpreted as *exactly* the quantity specified rather than as a lower or upper bound (or for that matter, an approximate quantity). Few studies in decision science have collected data to resolve this clash of assumptions. The present research provided such empirical tests and found support for the idea that people often interpret quantifiers in ways other than as "exactly that quantity."

In the present research, we hypothesized that people tend to interpret *half* as a lower bound—that is, tend to *lower bound* the quantifier—and that this would have predictable effects on their frequency estimation. The predicted effects were observed such that our manipulations of frame and probability level interacted as expected in each of three experiments. Support for the lower-bounding hypothesis was not only observed in the analyses of central tendencies, but also in the analyses of proportions, which showed that a majority of participants in each experiment provided frequency estimates that were consistent with lower bounding (even if responses of 500 also consistent with an *exactly* reading were removed), whereas only a minority gave estimates consistent with an *exactly* interpretation of *half*. Finally, contrary to the cognitive deficiency hypothesis, the tendency to lower bound the quantifier was unrelated to incoherence in Experiments 1 and 2, even when participants who responded with a value of 500 (i.e., exactly half) were excluded.

These findings generalize earlier research on the linguistic interpretation of quantifiers in judgment and decision-making tasks in various ways. First, whereas earlier work demonstrating a tendency to lower bound quantifiers in judgment and decision-making tasks focused on numeric quantifiers (Mandel 2014; Teigen and Nikolaisen 2009), the present research examined the tendency of participants to interpret a commonly used non-numeric quantifier (*half*) as a lower bound. Second, whereas Mandel (2014) examined the issue in the context of risky choice, the present experiments examined the issue in the context of frequency estimation. Third, whereas the numeric quantifiers studied in Mandel (2014) were not coupled with probabilities but rather expressed as explicit certainties (e.g., "it is certain that 200 people will be saved"), the present experiments examined how probability levels might be expected to interact with frames if participants interpreted the quantifier as a lower bound. Finally, whereas the values of the numeric quantifiers differed across frames in earlier work (i.e., "*200* people will be saved" was used in the gain frame, and "*400* people will die" was used in the loss frame), the quantifier *half* was held constant across frame in the present research (e.g., "half survive" vs. "half die").

The present findings are of methodological and theoretical significance because they demonstrate that quantifiers are not linguistically straightjacketed such that they must be read as meaning "exactly that much." Although Mandel (2014, Experiment 3) found that most participants, when given the standard format of the ADP, adopted a lower-bound interpretation of the numeric quantifiers presented in the sure options, the subsample that received the standard format of the ADP in that experiment was comparatively small, comprising only 50 participants. In contrast, the present research found strong evidence for a modal lower-bound-consistent frequency response based on data collected from 2088 participants. The implications of such findings for framing research, in particular, should be evident. If it is required that participants interpret the quantifiers in alternative frames as meaning "exactly that quantity," and if participants do not, in fact, interpret the quantifiers as intended, perhaps treating them instead as lower bounds, then the experiment cannot test what it purports to test. Convenient as it may be, researchers are not at liberty to decree by fiat that two statements are equivalent if people reading the statements disagree. Yet this is precisely what has happened in framing research for several decades (Fisher and Mandel 2021; Mandel 2022; Mandel and Vartanian 2011; Teigen 2011).

The failure to seriously consider the requirements of a framing task and to ensure that they are met has set back theoretical progress in this research area. Until the weaknesses of the experimental methods used to test for violations of description invariance are acknowledged, there will be a continuation of unclear thinking about the subject matter. For instance, it is not widely acknowledged that the requirements for framing are unlikely to be met in tasks such as the Asian Disease Problem and that they must be met in order for the logic of the experiment to follow through. Accordingly, the empirical effects of such experiments will continue to be incorrectly described as framing effects. This is why the term *framing* in the title of this article was set in scare quotes. If "half survive" and "half die" are interpreted as "at least half survive" and "at least half die," then they are no longer frames in a strict sense (Frisch 1993; Kühberger 1998; Mandel 2001, 2008, 2022, 2023). The replicability of the behavioral effect, moreover, which is not in question (e.g., Klein et al. 2014; Kühberger 1998), does nothing to resolve the issue, and pointing to it merely obfuscates the theoretical and methodological problems that we describe.

We do not deny that taking individual differences in linguistic interpretation of frames into account complicates methodological matters. Channeling Nagel (1986), it is far easier to attempt to impose a "view from nowhere" representation on the participant, but this is as self-defeating as it is delusional. It is self-defeating because if the experimental aim is to study how coherently individuals respond to alternative descriptions that they themselves regard as extensionally equivalent, then experimenters must collect data which sheds light on how the participant understands the verbal matter that forms the basis of the task. Or else, they must take precautions to ensure that the meaning that they intend to convey is what ought to be conveyed. For instance, if a task requires exactly some quantity to be specified, then it could be explicitly stated and even explicitly stated in different ways such as "exactly *n*" and "no more or less than *n*" (e.g., see Mandel 2008). To do otherwise, however—namely, to assume that the participant's representation just happens to map onto the experimenter's—is to commit what William James called the *psychologist's fallacy par excellence*. As James put it, "The *great* snare of the psychologist is the *confusion of his own standpoint with that of the mental fact* about which he is making his report (1890/1950, p. 196, italics in original). The current unverified approach is delusional because there is no legitimate view from nowhere. There is no "easily verified" objective meaning to statements such as "200 lives will be saved" or "half will survive", even if there happens to be strong intersubjective agreement on this issue among a group of researchers. The methodological implications of this research, of course, do not imply that variations in frame are impotent. It is simply that much less of what is called framing strictly qualifies as framing and, therefore, much that has been claimed about human rationality based on studies of strict framing is far more questionable than it is often assumed to be.

The present research is also clearly not a call to replace an exact reading of quantifier with a lower-bound reading. Rather, it is a call to reject any straightjacketed linguistic assumptions that are untested but which may, ultimately, undermine the logic of experiments in judgment and decision-making that have implications for theory, practice, and the broader characterization of human rationality. While the present work confirms the existence hypothesis that quantifiers can be interpreted in ways other than "exactly that quantity," we suspect that future studies could articulate a much fuller account of the conditions under which various types of quantifier interpretation are more likely to be manifested. For instance, in some contexts, it may be viewed as a more egregious error to overestimate a quantity than to underestimate it (Teigen and Filkuková 2011), and this might prompt lower-bound interpretations if receivers infer that senders are intentionally playing it safe with their estimates. However, we obtained evidence of lower bounding in the birth scenario of Experiment 3, which is unlikely to invite such conversational inferences, suggesting that this is not an overriding factor. One might also expect an attenuation of lower bounding if quantifiers do not convey round numbers such as those used in the present studies (Ferson et al. 2015; Sadock 1977).

The findings of Experiment 3 clearly indicate that individuals sometimes adopt linguistic interpretations of quantifiers that are, on the one hand, fuzzier than both single-bounded or exact readings, and on the other hand, conveying more than "roughly *n*" or "about *n*" would seem to convey. Specifically, whereas "at least *n*" implies the certainty of *n* but possibly more (Geurts and Nouwen 2007), it seems unlikely that in many cases in which quantifiers are used in uncertain assessments (such as probabilistic forecasts), receivers will be absolutely certain about the lower bound. In such instances, a receiver might instead interpret the sender's use of *half* to mean something like "*roughly* half, but if not *exactly* half, then more likely *more* than half rather than *less* than half." Future research could test this hypothesis, for instance, by assessing the probabilities assigned to various possible outcomes such as "less than half," "exactly half," and "more than half." If the quantifier is interpreted as *at least* that much, then the probability assigned to "less than half" should be zero. However, if the quantifier is interpreted fuzzily as *roughly half but more likely more than half than less than half*, we might expect a nonzero probability to be assigned to "less than half," and, further, we might expect that this probability would be smaller than the probability assigned to "more than half." Given that judgment and decision-making are often couched in uncertainty and ambiguity, a better theoretical understanding of approximators, including complex linguistic ensembles in which approximators are qualified by probabilities, seems vital for the design of coherent experiments and theories in the field.

Our findings also reveal that receivers' understanding of probabilistic assessments involving quantifiers is shaped by multiple attributes of the probabilities. For instance, across all studies, support for the lower-bounding hypothesis was stronger if probabilities were expressed in numeric rather than verbal terms. This is consistent with the view mutually expressed by researchers and participants that the meaning of probabilities as *degrees of probability* is more clearly conveyed by the numeric format than the verbal format (see Dhami and Mandel 2022, for a review). Note that, in contrast, verbal probabilities are regarded as more likely to leak information about senders' recommendations or preferences (Collins and Mandel 2019). The fact that the lower-bounding response pattern was weaker in the verbal condition, therefore, suggests that the effect is not primarily due to pragmatic inferences about the senders' preferred messaging, although we do not rule out such "information leakage" effects on judgment or choice (Sher and McKenzie 2006). More generally, our findings highlight that simply using numeric probabilities in assessments is insufficient to eliminate linguistic ambiguity and that the challenge for expert assessors is even greater than often suggested (e.g., Dhami and Mandel 2021; Irwin and Mandel 2023).

The lower-bound response pattern was also stronger when assessments were qualified by low rather than high probabilities. Although we did not anticipate this effect, we speculate that it may be the result of the marked nature of the low probability assessments: affirming the likelihood of "half …" may appear less pointed than negating that likelihood. Thus, the rejection of "half or more …" might have yielded a stronger lower-bounding response than the "mere" acceptance of "half or more …" Future studies could remove probabilities altogether and test this hypothesis by constructing assessments that categorically accept or reject prospects of ambiguously quantified outcomes. As well, future research might assess whether the assessments that include low probabilities are judged to be stronger, more forceful statements than those that include high probabilities, all else being equal. Such work may contribute to psycholinguistic theory, and it is certainly vital to putting judgment and decision research on a sounder methodological and theoretical footing.


### Acknowledgments

We thank Daniel Irwin, Nicole Herz, and Sarah Gibbon for their research assistance; David Budescu and Robert Collins for feedback on data visualizations; and Fintan Costello, Karl Teigen, and one anonymous reviewer for feedback on earlier drafts of this article.


### Disclosure

Portions of this research were presented at the Psychonomic Society 63rd Annual Meeting and the 2023 Subjective Probability, Utility, and Decision-Making Conference. This research contributes to the North Atlantic Treaty Organization Systems Analysis and Studies Panel Research Task Group on Anticipatory Intelligence for Superior Decision-Making (SAS-189).

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The data that support the findings of this study are openly available in the Open Science Framework at https://osf.io/m8kst/.

### References


Baron, J., S. Scott, K. Fincher, and S. E. Metz. 2015. "Why Does The Cognitive Reflection Test (Sometimes) Predict Utilitarian Moral Judgment (and Other Things)?" *Journal of Applied Research in Memory and Cognition* 4, no. 3: 265–284. https://doi.org/10.1016/j.jarmac.2014.09.003.

Bilker, W. B., M. R. Wierzbicki, C. M. Brensinger, R. E. Gur, and R. C. Gur. 2014. "Development of Abbreviated Eight-Item Form of The Penn Verbal Reasoning Test." *Assessment* 21, no. 6: 669–678. https://doi.org/10.1177/1073191114524270.

Borg, E. 2022. "IX—In Defence of Individual Rationality." *Proceedings of The Aristotelian Society* 122, no. 3: 195–217. https://doi.org/10.1093/arisoc/aoac009.

Breheny, R. 2008. "A New Look at The Semantics and Pragmatics of Numerically Quantified Noun Phrases." *Journal of Semantics* 25, no. 2: 93–139. https://doi.org/10.1093/jos/ffm016.

Carston, R. 1998. "Informativeness, Relevance and Scalar Implicature." In *Relevance Theory: Applications and Implications*, edited by R. Carston and S. Uchida, 179–236. Amsterdam, Netherlands: Benjamins.

Chierchia, G., D. Fox, and B. Spector. 2012. "Scalar Implicature as a Grammatical Phenomenon." In *Semantics: An International Handbook of Natural Language Meaning*, edited by C. Maienborn, K. von Heusinger, and P. Portner, vol. 3, 2297–2331. Berlin, Germany: Mouton de Gruyter.

Cokely, E. T., M. Galesic, E. Schulz, S. Ghazal, and R. Garcia-Retamero. 2012. "Measuring Risk Literacy: The Berlin Numeracy Test." *Judgment and Decision Making* 7, no. 1: 25–47. https://doi.org/10.1017/S1930297500001819.

Collins, R. N., and D. R. Mandel. 2019. "Cultivating Credibility With Probability Words and Numbers." *Judgment and Decision Making* 14, no. 6: 683–695. https://doi.org/10.1017/S1930297500005404.

Collins, R. N., D. R. Mandel, and B. Macleod. 2024. "Verbal and Numeric Probability Information Differentially Shapes Decisions." *Thinking & Reasoning* 30, no. 1: 235–257. https://doi.org/10.1080/13546783.2023.2220971.

Dhami, M. K., and D. R. Mandel. 2021. "Words or Numbers? Communicating Probability in Intelligence Analysis." *American Psychologist* 76, no. 3: 549–560. https://doi.org/10.1037/amp0000637.

Dhami, M. K., and D. R. Mandel. 2022. "Communicating Uncertainty Using Words and Numbers." *Trends in Cognitive Sciences* 26, no. 6: 514–526. https://doi.org/10.1016/j.tics.2022.03.002.

Faul, F., E. Erdfelder, A.-G. Lang, and A. Buchner. 2007. "G*Power 3: A Flexible Statistical Power Analysis Program for The Social, Behavioral, and Biomedical Sciences." *Behavior Research Methods* 39: 175–191. https://doi.org/10.3758/BF03193146.

Ferson, S., J. O'Rawe, A. Antonenko, et al. 2015. "Natural Language of Uncertainty: Numeric Hedge Words." *International Journal of Approximate Reasoning* 57: 19–39. https://doi.org/10.1016/j.ijar.2014.11.003.

Fisher, S., and D. R. Mandel. 2021. "Risky-Choice Framing and Rational Decision-Making." *Philosophy Compass* 16, no. 8: e12763. https://doi.org/10.1111/phc3.12763.

Frisch, D. 1993. "Reasons for Framing Effects." *Organizational Behavior and Human Decision Processes* 54, no. 3: 399–429. https://doi.org/10.1006/obhd.1993.1017.

Geurts, B. 2006. "Take "Five": The Meaning and Use of a Number Word." In *Non-Definiteness and Plurality*, edited by S. Vogeleer and L. Tasmowski, 311–329. Amsterdam, Netherlands: Benjamins.

Geurts, B., and R. Nouwen. 2007. "*At Least* et al.: The Semantics of Scalar Modifiers." *Language* 83, no. 3: 533–559. http://www.jstor.org/stable/40070901.

Haran, U., I. Ritov, and B. A. Mellers. 2013. "The Role of Actively Openminded Thinking in Information Acquisition, Accuracy, and Calibration." *Judgment and Decision Making* 8, no. 3: 188–201. https://doi.org/10.1017/S1930297500005921.

Ho, E., D. V. Budescu, M. K. Dhami, and D. R. Mandel. 2015. "Improving The Communication of Uncertainty in Climate Science and Intelligence

Analysis." *Behavioral Science & Policy* 1, no. 2: 43–55. https://doi.org/10.1353/bsp.2015.0015.

Horn, L. R. 1992. "The Said and The Unsaid." In *Salt II: Proceedings From The Second Conference on Semantics and Linguistic Theory*, edited by C. Barker and D. Dowty, 163–192. Columbus, Ohio: Ohio State University, Department of Linguistics.

Irwin, D., and D. R. Mandel. 2023. "Communicating Uncertainty in National Security Intelligence: Expert and Non-Expert Interpretations of and Preferences for Verbal and Numeric Formats." *Risk Analysis* 43: 943–957. https://doi.org/10.1111/risa.14009.

James, W. 1950. *The Principles of Psychology: Volumes 1 and 2*. New York, NY: Dover (Original work published 1890).

Klein, R. A., K. A. Ratliff, M. Vianello, et al. 2014. "Investigating Variation in Replicability: A "Many Labs" Replication Project." *Social Psychology* 45, no. 3: 142–152. https://doi.org/10.1027/1864-9335/a000178.

Kühberger, A. 1998. "The Influence of Framing on Risky Decisions: A Meta-Analysis." *Organizational Behavior and Human Decision Processes* 75, no. 1: 23–55. https://doi.org/10.1006/obhd.1998.2781.

Levinson, S. C. 2000. *Presumptive Meanings*. Cambridge, MA: MIT Press.

Lipkus, I. M., G. Samsa, and B. K. Rimer. 2001. "General Performance on a Numeracy Scale Among Highly Educated Samples." *Medical Decision Making* 21, no. 1: 37–44. https://doi.org/10.1177/0272989X0102100105.

Mandel, D. R. 2001. "Gain-Loss Framing and Choice: Separating Outcome Formulations From Descriptor Formulations." *Organizational Behavior and Human Decision Processes* 85, no. 1: 56–76. https://doi.org/10.1006/obhd.2000.2932.

Mandel, D. R. 2008. "Violations of Coherence in Subjective Probability: A Representational and Assessment Processes Account." *Cognition* 106, no. 1: 130–156. https://doi.org/10.1016/j.cognition.2007.01.001.

Mandel, D. R. 2014. "Do Framing Effects Reveal Irrational Choice?" *Journal of Experimental Psychology: General* 143, no. 3: 1185–1198. https://doi.org/10.1037/a0034207.

Mandel, D. R. 2015a. "Accuracy of Intelligence Forecasts From The Intelligence Consumer's Perspective." *Policy Insights From The Behavioral and Brain Sciences* 2, no. 1: 111–120. https://doi.org/10.1177/2372732215602907.

Mandel, D. R. 2015b. "Communicating Numeric Quantities in Context: Implications for Decision Science and Rationality Claims." *Frontiers in Psychology* 6: 537. https://doi.org/10.3389/fpsyg.2015.00537.

Mandel, D. R. 2022. "Framing, Equivalence, and Rational Inference." *Behavioral and Brain Sciences* 45: e234. https://doi.org/10.1017/S0140525X22000954.

Mandel, D. R. 2023. "Subjective Equivalence: A Basic Requirement for Strict Framing Effects: Commentary on Huizenga et al. (2023)." *Decision* 10, no. 3: 235–237. https://doi.org/10.1037/dec0000206.

Mandel, D. R., M. K. Dhami, S. Tran, and D. Irwin. 2021. "Arithmetic Computation With Probability Words and Numbers." *Journal of Behavioral Decision Making* 34, no. 4: 593–608. https://doi.org/10.1002/bdm.2232.

Mandel, D. R., and D. Irwin. 2021. "Facilitating Sender-Receiver Agreement in Communicated Probabilities: Is It Best to Use Words, Numbers or Both?" *Judgment and Decision Making* 16, no. 2: 363–393. https://doi.org/10.1017/S1930297500008603.

Mandel, D. R., and I. Kapler. 2018. "Cognitive Style and Frame Susceptibility in Decision-Making." *Frontiers in Psychology* 9: 1461. https://doi.org/10.3389/fpsyg.2018.01461.

Mandel, D. R., and O. Vartanian. 2011. "Frames, Brains, and Content Domains: Neural and Behavioral Effects of Descriptive Content on Preferential Choice." In *Neuroscience of Decision Making*, edited by O. Vartanian and D. R. Mandel, 45–70. New York, NY: Psychology Press.

Mellers, B. A., J. D. Baker, E. Chen, D. R. Mandel, and P. E. Tetlock. 2017. "How Generalizable Is Good Judgment? A Multi-Task, Multi-Benchmark Study." *Judgment and Decision Making* 12, no. 4: 369–381. https://doi.org/10.1017/S1930297500006240.

Nagel, T. 1986. *The View From Nowhere*. Oxford, UK: Oxford University Press.

Oppenheimer, D. M., T. Meyvis, and N. Davidenko. 2009. "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power." *Journal of Experimental Social Psychology* 45, no. 4: 867–872. https://doi.org/10.1016/j.jesp.2009.03.009.

Sadock, J. M. 1977. "Truth and Approximations." *Proceedings of the 3rd Annual Meeting of the Berkeley Linguistics Society*, 430-439. https://doi.org/10.3765/bls.v3i0.2268.

Shafir, E., and R. A. LeBoeuf. 2002. "Rationality." *Annual Review of Psychology* 53: 491–517. https://doi.org/10.1146/annurev.psych.53.100901.135213.

Sher, S., and C. R. M. McKenzie. 2006. "Information Leakage From Logically Equivalent Frames." *Cognition* 101, no. 3: 467–494. https://doi.org/10.1016/j.cognition.2005.11.001.

Smithson, M., D. V. Budescu, S. B. Broomell, and H.-H. Por. 2012. "Never Say "Not": Impact of Negative Wording in Probability Phrases on Imprecise Probability Judgments." *International Journal of Approximate Reasoning* 53, no. 8: 1262–1270. https://doi.org/10.1016/j.ijar.2012.06.019.

Spector, B. 2013. "Bare Numerals and Scalar Implicatures." *Language and Linguistics Compass* 7, no. 5: 273–294. https://doi.org/10.1111/lnc3.12018.

Teigen, K. H. 2011. "When Frames Meet Realities: On The Perceived Correctness of Inaccurate Estimates." In *Perspectives on Framing*, edited by G. Keren, 197–217. New York, NY: Psychology Press.

Teigen, K. H., and W. Brun. 1995. "Yes, but It Is Uncertain: Direction and Communicative Intention of Verbal Probabilistic Terms." *Acta Psychologica* 88, no. 3: 233–258. https://doi.org/10.1016/0001-6918(93)E0071-9.

Teigen, K. H., and P. Filkuková. 2011. "Are Lies More Wrong Than Errors? Accuracy Judgments of Inaccurate Statements." *Scandinavian Journal of Psychology* 52: 8–20. https://doi.org/10.1111/j.1467-9450.2010.00843.x.

Teigen, K. H., and M. I. Nikolaisen. 2009. "Incorrect Estimates and False Reports: How Framing Modifies Truth." *Thinking & Reasoning* 15, no. 3: 268–293. https://doi.org/10.1080/13546780903020999.

Toplak, M. E., R. F. West, and K. E. Stanovich. 2017. "Real-World Correlates of Performance on Heuristics and Biases Tasks in a Community Sample: Heuristics and Biases Tasks and Outcomes." *Journal of Behavioral Decision Making* 30, no. 2: 541–554. https://doi.org/10.1002/bdm.1973.

Tversky, A., and D. Kahneman. 1981. "The Framing of Decisions and The Psychology of Choice." *Science* 211: 453–458. https://doi.org/10.1126/science.7455683.

van Rooij, R. 2006. "Pragmatic Meaning and Non-Monotonic Reasoning: The Case of Exhaustive Interpretation." *Linguistics and Philosophy* 29: 205–250. https://doi.org/10.1007/s10988-005-3760-4.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.