An experimental investigation of the *around* / *between* contrast

Adèle Hénot-Mortier, Steven Verheyen, Paul Égré, Benjamin Spector

Abstract

A recent Bayesian account of numerical approximation expressions (Égré et al., 2023) makes the central prediction that the posterior distribution over possible numbers induced by an expression of the form *around n* should be more peaked around its central value *n* than a posterior induced by an equivalent expression of the form *between x and y*. In this paper, we test and verify this prediction using a novel online probability-elicitation paradigm. Our result confirms the finding that inherently vague expressions like *around n* convey distinctive information about what the distribution of possible numbers is according to the speaker, in a way that uncertain (but not vague) expressions like *between x and y* do not.

1 Around vs. between

1.1 An initial puzzle

Modifiers such as *around*, *roughly*, *approximately*, are called "approximators" by Sauerland and Stateva (2007). When applied to numerals, approximators end up denoting a wider range of possible values than the numeral they modify. For instance, (1a) could be uttered if 17, 23, or 20 people came to the party. More specifically, approximated numerals seem to denote an interval which contains the numeral itself, and which is further constrained by contextual and pragmatic considerations (Krifka, 2007; Solt, 2015). Solt (2015) for instance, proposes that expressions such as *around n* denote an interval centered around n, and bounded by the midpoints between two sequential elements on some specific contextually-provided granularity scale S_G (e.g., if the granularity is $G = 10$, $S_G = [0, 10, 20, 30...]$).

Uncertain numerical expressions, such as *between 15 and 25* in (1b), also denote an interval. The bounds of this interval, unlike the ones of an interval evoked by *around n*, are fully precise and not context-dependent. (1b) will not be felicitous if 14 or 26 people came to the party, but will be if 15 or 25 people came. In other words, *between 15 and 25* straightforwardly denotes the interval $[15; 25]$. This is generalized in (2b).

(1) *How many people came to the party?*
  a. **Around** 20 people came to the party.
  b. **Between** 15 and 25 people came to the party.

(2) a. $[\![\text{ around n }]\!]^g = [n - {}^G/_2; n + {}^G/_2]$
  b. $[\![\text{ between x and y }]\!]^g = [x; y]$

In sum, even if both approximated numerals (like *around 20*) and uncertain numerals (like *between 15 and 25*) express uncertainty about an exact numerical value k, they are not totally synonymous. *Between x and y* conveys uncertainty within a fully *precise* domain (namely, the interval $[x; y]$), while *around n* conveys uncertainty within an underspecified domain. The question is then:

why is *around n* used at all, if, for each value of n, it is possible to find an expression of the form *between x and y* that precisely and overtly denotes the intended interval?

1.2 A Bayesian solution

A solution to the above question, which also does more justice to the inherent vagueness conveyed by *around*, stems from two observations. The first, is that *around n* is Sorites-susceptible (Lassiter and Goodman, 2015; Égré et al., 2019): if k can be taken to be *around n*, then so do $k-1$ and $k+1$. This cannot be captured if we assume that *around n* denotes one sharp interval, even if it is taken to be context-dependent. The second observation is that *around n*, unlike *between x and y*, does not seem to associate each possible numerical value with same likelihood (Égré, 2022; Égré et al., 2023): upon hearing *around n*, values closer to n appear to be more likely than values remote from n. These two observations motivate a probabilistic account of approximated and uncertain numerals, which assigns *around n* and *between x and y* distinct probability distributions over possible numerical values.

Égré et al. (2023) provide such an account within a Bayesian framework, by devising probabilistic update rules for *around n* and *between x and y*. This account comes with one key prediction, which can be stated informally as a "peakedness" contrast: given two expressions of the form *around n* and *between x and y*, respectively yielding the posteriors $\mathbb{P}_a$ and $\mathbb{P}_b$, if $\mathbb{P}_a$ and $\mathbb{P}_b$ share the same support, then, regardless on what the prior distribution in possible numbers was, $\mathbb{P}_a$ will be more peaked around its central value than $\mathbb{P}_b$. In this paper, we propose to test the empirical validity of this prediction via a series of online posterior-elicitation experiments. In the next section we sketch Égré et al.'s account and its prediction. Section 3 will de dedicated to the design of the experiments and section 4 to the results. Section 5 concludes.

2 A Bayesian account of *around* and *between* and its predictions

In the spirit of previous Bayesian accounts of pragmatic inference, (Lassiter and Goodman, 2013; Bergen et al., 2016; Qing and Franke, 2015), Égré et al. (2023) assume that pragmatic meaning arises from the interaction between two Bayesian reasoners: a speaker $\mathcal{S}$, and a listener $\mathcal{L}$. $\mathcal{S}$ is assumed to produce an expression of the form *around n* or *between x and y*, while $\mathcal{L}$ is assumed to compute a posterior distribution over possible numerical values, given what $\mathcal{S}$ said. The computation of this posterior follows Bayes' rule (3), where k is a random variable ranging over possible numerical values, and u is a random variable ranging over possible utterances (in our case, of the form *around n* or *between x and y*). The posterior associated with *around n* or *between x and y* then directly depends, for each k, on the likelihood that either expression was used to convey k.

(3)
$$\underbrace{\mathbb{P}[k|u]}_{\substack{\text{posterior over number } k\\ \text{given utterance } u}} = \underbrace{\mathbb{P}[k]}_{\text{prior over number } k} \times \frac{\overbrace{\mathbb{P}[u|k]}^{\text{likelihood of } u \text{ given } k}}{\underbrace{\mathbb{P}[u]}_{\text{marginal over } u}}$$

2.1 *Between*

If $\mathcal{S}$ utters $u =$ *between x and y*, the likelihood of u given k $\mathbb{P}[u|k]$ will be non-zero iff $k \in [x; y]$. Moreover, it is assumed to be uniform on this interval. This yields the update rule in (4). This update crucially keeps the *relative* probabilities of the values within the interval as they were according to the prior. In summary, the utterance of *between x and y* modifies the prior distribution over possible numerical values by truncating its support to fit the interval $[x; y]$, and renormalizing the result.

(4) $$\mathbb{P}[k|\textit{between x and y}] \propto \begin{cases} \mathbb{P}[k] & \text{if} \quad k \in [x; y] \\ 0 & \text{if} \quad k \notin [x; y] \end{cases}$$

2.2 *Around*

Building on Solt (2015), Égré et al. (2023) assume the lexical entry in (5) for *around*. The major difference between this entry and the one in (2a) is that i in (5) is taken to be a free, underspecified variable, while G in (2a) is a variable that is *set* by the context and tied to granularity considerations. In other words, the entry in (5) is vague in a way that the one in (2a) is not.

(5) $$[\![\text{ around n }]\!]^i = \lambda x.\ x \in [n - i; n + i]$$

If $\mathcal{S}$ utters $u =$ *around n*, Égré et al. (2023) assume that the listener $\mathcal{L}$ draws inferences about i (half-width of the interval) and k (numerical value of interest) *at the same time*. Moreover, it is assumed that the listener's joint prior probability distribution $\mathbb{P}[k, i]$, is such that k and i are probabilistically independent, which means that it can be decomposed into the product of the individual priors (6). This independence assumption is motivated by the fact that the value of k corresponds to a real-world observation, while i is a parameter associated with language use, and as such *a priori* independent of any non-linguistic fact.

(6) $$\mathbb{P}[k, i] = \mathbb{P}[k] \times \mathbb{P}[i]$$

The posterior induced by *around n* can then be computed from (7) using the Law of total probability over the set of all possible values of i

(7) $$\begin{aligned} \mathbb{P}[k|\textit{around n}] &= \textstyle\sum_{i=0}^{n} \mathbb{P}[i] \times \mathbb{P}[k|\textit{around } n, i] \\ &= \textstyle\sum_{i=0}^{n} \mathbb{P}[i] \times \mathbb{P}[k|\textit{between n-i and n+i}] \\ &\propto \textstyle\sum_{i=0}^{n} \mathbb{P}[i] \times \mathbb{P}[k] \times \mathbb{1}_{k \in [n-i; n+i]} \\ &\propto \mathbb{P}[k] \times \textstyle\sum_{i=|n-k|}^{n} \mathbb{P}[i] \end{aligned}$$

The presence of a sum whose total number of terms $(n - |n - k| + 1)$ *decreases* as the distance between n and k ($|n - k|$) *increases* in the above update rule translates into the following fact: the closer k is to n, the higher the posterior assigned to k given *u=around n*. This itself stems from the observation that a value of k close to n belongs to *more* intervals centered around n, than a value of k remote from n does (as schematized in Figure (1)). As a result of Bayesian reasoning then, the posterior probabilities assigned to numbers closer to n are predicted to be greater than their prior probabilities, while numbers more remote from n will receive lower posterior probabilities than assigned by the prior distribution.

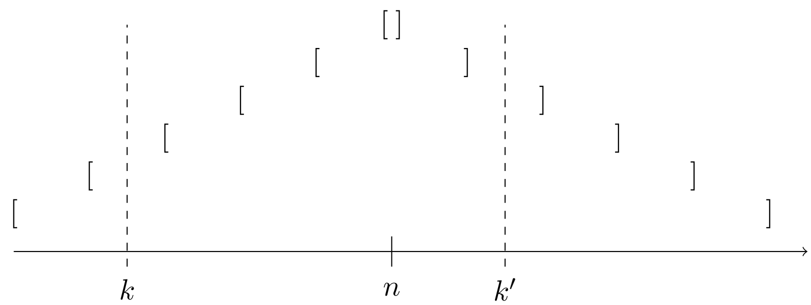


Figure 1: k' (closer to n) belongs to 4 intervals centered around n, while k (further from n) only belongs to 2 such intervals.

2.3 Key predictions

A direct consequence of the Bayesian updates related to *around n* and *between x and y* is the following: given two possible numerical values k and k', such that k is closer to n than k' is, and two posterior distributions $\mathbb{P}[k|\textit{around n}]$ and $\mathbb{P}[k|\textit{between x and y}]$ sharing the same support S (with $k, k' \in S$), the ratio of the posterior probabilities of k and k' is predicted to be greater for the former distribution (i.e. after updating with an *around n* statement) than for the later distribution (i.e. after updating with a *between x and y* statement).

(8) $$\frac{\mathbb{P}[k|\textit{around n}]}{\mathbb{P}[k'|\textit{around n}]} \geq \frac{\mathbb{P}[k|\textit{between x and y}]}{\mathbb{P}[k'|\textit{between x and y}]}$$

The above inequality is true *regardless of the particular prior distribution over* k, which makes it possible to test on different participants, who by default may not have the same priors. However, given two distributions $\mathbb{P}[k|\textit{around n}]$ and $\mathbb{P}[k|\textit{between x and y}]$ sharing the same support $S = [n-i; n+i]$, there will be $4 \times \sum_{\epsilon=1}^{i} i - \epsilon = 2i(i-1)$ possible pairs (k, k') verifying the preconditions of this inequality. A more empirically robust prediction can be obtained by averaging each side of the inequality (8) over all possible pairs (k, k') verifying its preconditions:

(9) $$\frac{1}{|K|} \sum_{(k,k') \in K} \frac{\mathbb{P}[k|\textit{around n}]}{\mathbb{P}[k'|\textit{around n}]} \geq \frac{1}{|K|} \sum_{(k,k') \in K} \frac{\mathbb{P}[k|\textit{between x and y}]}{\mathbb{P}[k'|\textit{between x and y}]}$$

Where $K = \{(k, k') \in S \mid |n-k| < |n-k'|\}$.

The next section presents our experimental paradigm and how it addresses the above prediction.

3 Experimental design

3.1 General paradigm

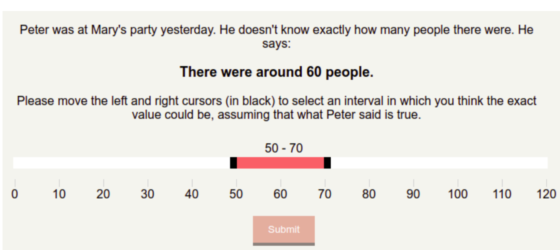
We designed a posterior-elicitation paradigm that could be carried on *via* a web interface. The goal was to elicit empirical posterior distributions corresponding to $\mathbb{P}[k|\textit{around n}]$ and $\mathbb{P}[k|\textit{between x and y}]$, in order to compute the relevant ratios and test (a slight adaptation of) inequality (9). To that end, participants were presented with a short (written) context, followed by a critical statement containing an expression of the form *around n* or *between x and y*. Participants were tested on both kinds of expressions, in order to assess inequality (9) within-participant. The main challenge, from an empirical perspective, was to ensure that the variables n, x, and y, were such that the *around n* and the *between x and y* expressions presented to each participant, verified the precondition that their posteriors share the same support.

Each participant was tested on one pair of expressions of the form *around n/between x and y*. Values of n were randomly set to 40, 50 or 60 for each participant. Values of x and y were dynamically determined for each specific participant, in a way that will be made clear in the next paragraphs, and was meant to ensure each participant was exposed to *around* and *between* expressions whose induced posteriors share the same support. The template used for the context and critical sentence is given in Figure 2.

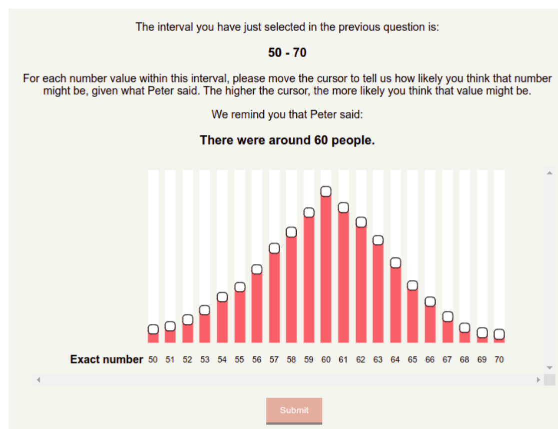
Peter was at Mary's party yesterday. He doesn't know exactly how many people there were. He says:
There were {around n | between x and y} people.

Figure 2: Template for the target sentence and its context.

A trial for a fixed *around n* or *between x and y* expression was comprised of two main tasks: an INTERVAL task (cf. Figure 3a) and a HISTOGRAM task (cf. Figure 3b). For each such expression, the INTERVAL task would always precede the HISTOGRAM task, and moreover the INTERVAL task related to the *around* expression would always precede that of the related *between* expression.



(a) INTERVAL task



(b) HISTOGRAM task

Figure 3: Screen captures of the two tasks (n randomly set to 60)

The goal of the INTERVAL task was twofold. First, it was used to elicit the support S related to the posterior of a given expression (*around* or *between*), according to each participant. In practice, participants were asked to move two sliders on a horizontal bar representing the $[0; 120]$ interval (see Figure 3a) to set the value of S. The two bounds were supposed to define the interval within which the exact numerical value should be, given the target sentence and its context. The second goal of this task was to determine the values of x and y to use in the *between* expression corresponding to each *around* expression, in order to have two expressions comparable by the means of inequality (9). For instance, if *around 40* gave rise to a support $S = [35; 45]$ according to a certain participant, the corresponding *between* expression presented to them would be *between 35 and 45*. In other words, the values of the lower and upper bounds returned by the participant during the INTERVAL task testing *around* were reused as respectively the x and y values of the *between* expression. Since the support of *between x and y* is supposed to be the interval $[x; y]$, this mechanism was meant to guarantee that the support of the posterior distributions induced by *around n* and *between x and y* were the same *for each individual participant*.

The Histogram task was designed to elicit the posterior distributions $\mathbb{P}[k|\textit{around } n]$ and $\mathbb{P}[k|\textit{between } x \textit{ and } y]$, defined on the supports returned by the participant during the Interval task. During this task, the participant was again presented with the short context and critical sentence, and was reminded of the interval returned in the preceding Interval task. The participant was then asked to "draw" a posterior distribution in the form of a histogram, by using vertical sliders, one per possible number within the support (see Figure 3b). The resulting histogram was intended to reflect the likelihood of each possible value within the support, given the critical approximation statement. The weights associated with each possible numerical value were collected and normalized.

3.2 Prediction tested

In the experiments described in the next section, a slight variant of inequality (9) was tested. First, the pairs (k, k') such that either k or k' is a "salient" value – n of *around* n, x, or y of *between x and y* – were excluded from the set of pairs over which individual posterior ratios were averaged (cf. inequality (10)). Those values were excluded because it is reasonable to think that them being part of the target expressions may increase the salience of alternative expressions featuring those numbers (e.g., (*exactly*) n), and in turn introduce additional pragmatic inferences regarding those particular numbers, that were not meant to be captured by Égré et al.'s model.[1]

(10) $$\frac{1}{|K|} \sum_{\substack{(k,k') \in K \\ k,k' \notin \{n,x,y\}}} \frac{\mathbb{P}[k|\textit{around } n]}{\mathbb{P}[k'|\textit{around } n]} \geq \frac{1}{|K|} \sum_{\substack{(k,k') \in K \\ k,k' \notin \{n,x,y\}}} \frac{\mathbb{P}[k|\textit{between } x \textit{ and } y]}{\mathbb{P}[k'|\textit{between } x \textit{ and } y]}$$

Additionally, differences in the supports S_a and S_b of the posteriors induced by respectively *around* and *between* – which were meant to be minimized by the paradigm presented in the previous section – were accommodated by considering pairs (k, k') in their intersection ($S_a \cap S_b$). This yields inequality (11), where K is replaced by $K' = \{(k, k') \in S_a \cap S_b \mid |n - k| < |n - k'|\}$.

(11) $$\underbrace{\frac{1}{|K|} \sum_{\substack{(k,k') \in K' \\ k,k' \notin \{n,x,y\}}} \frac{\mathbb{P}[k|\textit{around } n]}{\mathbb{P}[k'|\textit{around } n]}}_{R_a} \geq \underbrace{\frac{1}{|K|} \sum_{\substack{(k,k') \in K' \\ k,k' \notin \{n,x,y\}}} \frac{\mathbb{P}[k|\textit{between } x \textit{ and } y]}{\mathbb{P}[k'|\textit{between } x \textit{ and } y]}}_{R_b}$$

We call R_a (for "*around*-ratio") and R_b (for "*between*-ratio") the left-hand side and right-hand side of inequality (11), respectively. Inequality (11) is the one we propose to test in the series of experiments presented in section 4.

3.3 Participant recruiting and filtering

All the experiments presented in the next sections were preregistered on OSF. They were all conducted on US-based Amazon Mechanical Turk Master Workers, meaning, participants who have consistently demonstrated a high degree of success in performing a wide range of tasks

[1] As an example, the probability of n after updating with *around* n might be undermined by competition with (*exactly*) n, leading ratios involving $\mathbb{P}[n|\textit{around } n]$ in their numerator to be lower than initially predicted.

across a large number of Amazon MTurk Requesters. We also required that the participants have a success rate of at least 98% on previous MTurk tasks.

In all experiments, participants were excluded based on demographic factors and quality requirements. First, participants who signaled that they were not native speakers of English were excluded. The remaining participants who were 1.645 standard deviations faster than the average completion time were also excluded. Finally, participants who provided inconsistent intervals were excluded. Inconsistency was defined as non-overlap with $[x; y]$ in the case of *between x and y*; and non-overlap with the central value n in the case of *around n*. Note that this condition was relatively permissive, which meant that *around/between* pairs not sharing the exact same support were included.

4 Experiments

4.1 Initial Experiment (Experiment 1)

The first Experiment was conducted on 207 participants (determined *via* a power analysis based on pilot data). Each trial was comprised of an INTERVAL task immediately followed by the corresponding HISTOGRAM task. In addition to the two critical trials, this experiment involved six filler expressions featuring numbers with varying degrees of granularity. The target *around* and *between* items were presented at two fixed positions in the experiment, as 4th and 8th item respectively. The positions of the filler items were randomized over the set of remaining positions ($\{1, 2, 3, 5, 6, 7\}$). The choice to associate filler expression with different levels of granularity was intended to conceal the link between the critical *around n* expression (position 4) and the critical *between x and y* expression (position 8), whose x and y variables may be coarse- or fine-grained, depending on the participant's answers to the *around n* INTERVAL trial. It is also worth pointing out that because the INTERVAL and HISTOGRAM tasks related to a fixed expression were set to be adjacent within a trial, and because of the dependency between the x and y variables of the critical *between x and y* expression and the critical trial testing *around n*, the order of these two trials could not be randomized. Table 1 summarizes the design of this experiment.

Expression	Position	Category
around n	4	Critical
between x and y	8	Critical
between 80 and 90	random({1, 2, 3, 5, 6, 7})	Filler
between 13 and 24	random({1, 2, 3, 5, 6, 7})	Filler
around 70	random({1, 2, 3, 5, 6, 7})	Filler
around 36	random({1, 2, 3, 5, 6, 7})	Filler
almost 60	random({1, 2, 3, 5, 6, 7})	Filler
almost 24	random({1, 2, 3, 5, 6, 7})	Filler

Table 1: Trials for Experiment 1

A two-tailed Wilcoxon signed-rank test for matched pairs with a significance threshold of .05 was used to test the hypothesis $R_a \neq R_b$ within-participant, on a sample size of 162 (after all exclusions). It turned out significant in favor of $R_a > R_b$ ($p = 8.5$e-13), in line with inequation (11). This effect was associated with a medium effect size (Cohen's $d = .49$). Exploratory, Holm-Bonferroni-corrected two-tailed Wilcoxon tests performed on each subset of participants who were assigned to the same n also ended up significant ($p_{corr} = 7.2\text{e-}5$ for $n = 40$; $p_{corr} = 1.2\text{e-}6$ for $n = 50$; $p_{corr} = 1.0\text{e-}4$ for $n = 60$), and were associated with medium effect sizes ($d = .51$ for $n = 40$; $d = .57$ for $n = 50$; $d = .48$ for $n = 60$).

Group-by-group averaged results are given in Figure 4 and confirm the within-participant results, in that posteriors induced by *around* tend to be more peaked than posteriors induced by a similar *between* expression. It is however worth noting that the overall shape of those averaged results – both the curves and their envelopes – can be misleading. Because each participant defined their posterior distribution on potentially different intervals, which had the property to be approximately centered around the values $n = 40$, 50 or 60, the action of averaging posteriors *per se* is expected to boost the probability of the values that belong to more intervals across participants (typically, values close to n), and to undermine the probability of the values that belong to fewer intervals.[2] In other words, averaged posteriors tend to be more peaked than the individual posteriors used to compute them. The 95%-confidence envelopes on the other hand, might give the impression of a lower significance than actually observed. This is again because the envelopes are computed from averaged data, while the variation at stake was within-participant. The contrast between two averaged, paired distribution such as the *around* and *between* posteriors however, remains informative.

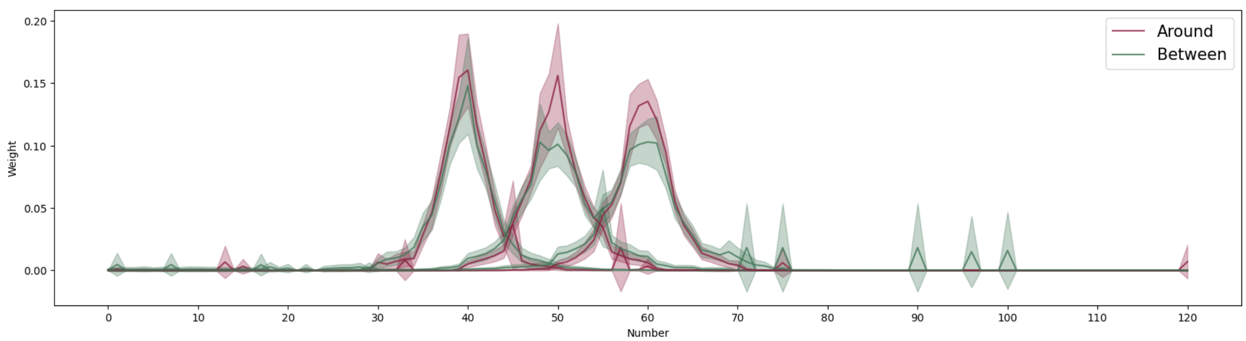


Figure 4: Averaged posteriors for $n = 40$, 50 or 60 in Experiment 1. Envelopes define 95% confidence intervals around the means.

A concerning observation however, was that participants tended to provide distributions that were more uniform as they advanced through the study, which constitutes a potential confound that we tried to address in two follow-up experiments (Experiment 2 and 3).

4.2 Follow-up: Experiment 2

In order to decrease the distance between the two critical trials and in turn reduce the "flattening" effect driven by the number of completed trials, Experiment 2 replicated Experiment 1 without the fillers. It was conducted on 150 participants, given the more precise exclusion rate derived from Experiment 1. The design of this experiment is given in Figure 2. Just like in Experiment 1, Interval and Histogram tasks related to the same expression were set to be adjacent.

[2]For instance, it can be shown analytically that averaging *uniform* distributions centered around n but defined on intervals of incrementally increasing length, gives rise to a distribution that is sharlpy peaked around n.

Expression	Position	Category
around n (INTERVAL+HISTOGRAM)	1	Critical
between x and y (INTERVAL+HISTOGRAM)	2	Critical

Table 2: Trials for Experiment 2

We used a one-tailed Sign test with a significance threshold of .05 to test the hypothesis $R_a \neq R_b$ within-participant, on a sample size of 122 (after all exclusions). It turned out significant in favor of $R_a > R_b$ ($p = 2.5$e-8), again in line with inequation (11). This effect was associated with a medium effect size (Cohen's $d = .64$). Holm-Bonferroni-corrected, one-tailed Wilcoxon tests performed on each subset of participants who were assigned to the same n also ended up significant, as well (for $n = 40$, $p_{corr} = 1.7$e-3; for $n = 50$, $p_{corr} = 1.3$e-3; for $n = 60$, $p_{corr} = 5.1$e-5). Corresponding effect sizes were $.54$ (medium) for $n = 40$; $.70$ (medium) for $n = 50$; 1.02 (large) for $n = 60$. Group-by-group averaged results are given in Figure 5 and confirm the within-participant results, although in a way that is slightly less clear than in Experiment 1.

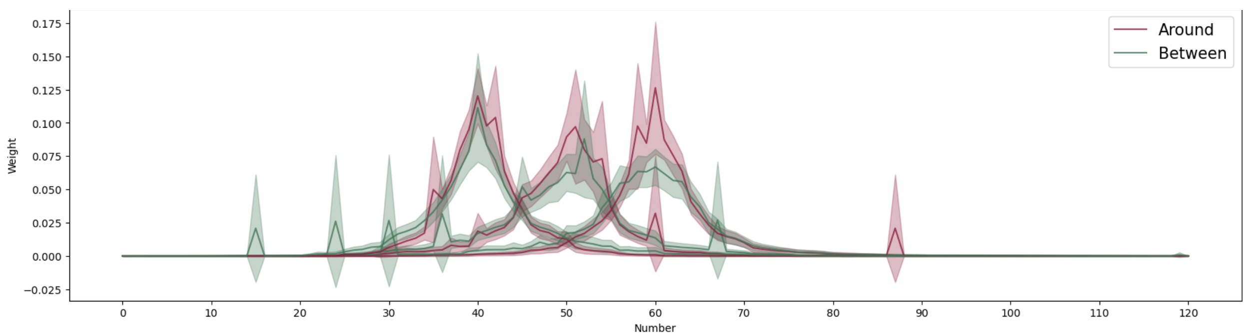


Figure 5: Averaged posteriors for $n = 40$, 50 or 60 in Experiment 2. Envelopes define 95% confidence intervals around the means.

Even though Experiment 2 corroborated the result of Experiment 1, its design did not allow to totally remove the potential order effect between *around* and *between*, since the HISTOGRAM task related to *around* still systematically preceded the HISTOGRAM task related to *between*. This issue is addressed in Experiment 3.

4.3 Follow-up: Experiment 3

Experiment 3 differs from Experiment 2 only in that the two HISTOGRAM tasks are fully randomized across participants. This is done by uncoupling those tasks from their corresponding INTERVAL tasks. More specifically, Experiment 3 started with two (non-randomized) INTERVAL tasks, testing first *around n*, and then *between x and y* (x and y still being dynamically determined). The two INTERVAL tasks were then followed by the corresponding HISTOGRAM tasks, in a randomized order. The design of this experiment is summarized below.

Expression	Position	Category
around n (Interval)	1	Critical
between x and y (Interval)	2	Critical
around n (Histogram)	random($\{3,4\}$)	Critical
between x and y (Histogram)	random($\{3,4\}$)	Critical

Table 3: Trials for Experiment 3

The target sample size for Experiment 3 was computed *via* a power analysis using as target effect size half the effect-size of Experiment 2; and was thus set to 250. The experiment was conducted on 286 participants, based on the exclusion rates from Experiments 1 and 2.

A one-tailed Sign test was performed on a final sample of 240 participants after all exclusions, and turned out significant ($p = 3.1$e-8). The associated effect size was small (Cohen's $d = .30$). Holm-Bonferroni-corrected one-tailed Wilcoxon tests performed on each subset of participants who were assigned to the same n also ended up significant ($p_{corr} = 3.0$e-5 for $n = 40$; $p_{corr} = 5.2$e-4 for $n = 50$; $p_{corr} = 4.2$e-4) for $n = 60$. Corresponding effect sizes were .28 (small) for $n = 40$; .34 (small/medium) for $n = 50$; .26 (large) for $n = 60$. Group-by-group averaged results are given in Figure 5 and confirm the within-participant results, though in a way that less clear than in both previous experiments.

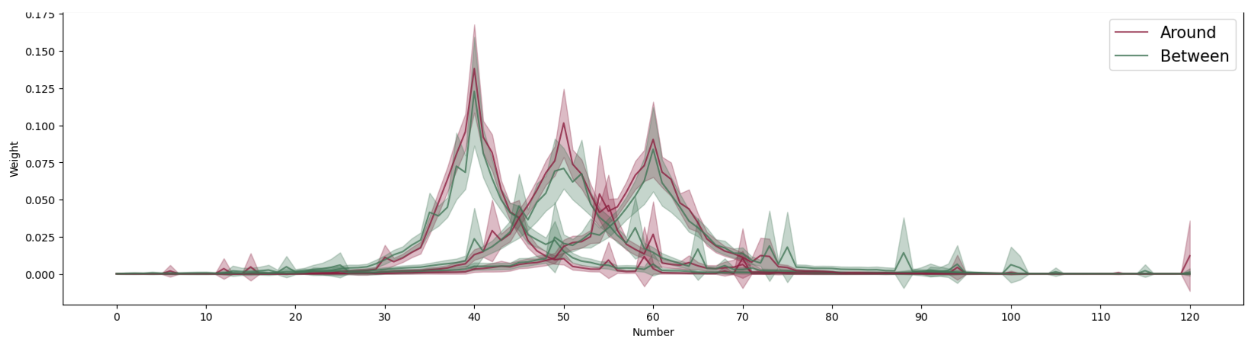


Figure 6: Averaged posteriors for $n = 40$, 50 or 60 in Experiment 3. Envelopes define 95% confidence intervals around the means.

In sum, the fully randomized design of Experiment 3 confirms the results of both previous experiments, though with smaller effect sizes, potentially due to the neutralization of confounding order effects.

Conclusion and discussion

The Bayesian model developed by Égré et al. (2023) proposes that the interpretation of an *around*-statement in a given context is inherently probabilistic. This account yields the prediction that statements containing *around n* can be used to communicate fine-grained probabilistic information in a way that alternative *between*-statements cannot. In particular, *around n* is predicted to induce a posterior distribution over possible numerical values that is more peaked than the distribution induced by a sufficiently similar *between x and y* expression, regardless of the prior distribution on possible numbers. The results of our series of studies appear robustly consistent this prediction and constitute the first empirical validation of the model. It is however worth keeping in mind that the prediction tested is a *consequence* of Égré et al.'s Bayesian model, and as such, could in principle be consistent with other accounts of *around* vs. *between*. To corroborate the result of this investigation, one may want to test a more direct prediction of Égré et al.'s model; for instance, by assessing inequalities between individual ratios-of-posteriors (using a single (k, k') pair) instead of between their means. One may also want to test if the Bayesian updates proposed for *around* and *between* are good models of the reasoning used by humans when they assess the probability of possible numbers upon hearing *around n* and *between x and y*. This kind of investigation however, would require to elicit both posterior and prior distributions.

References

Bergen, L., Levy, R., and Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9.

Channell, J. (1994). *Vague Language*. Oxford University Press.

Cummins, C., Sauerland, U., and Solt, S. (2012). Granularity and scalar implicature in numerical expressions. *Linguistics and Philosophy*, 35(2):135–169.

Frazee, J. and Beaver, D. (2010). *Vagueness Is Rational under Uncertainty*, page 153–162. Springer Berlin Heidelberg.

Kennedy, C. (2007). Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30(1):1–45.

Krifka, M. (2007). Approximate interpretation of number words. In Bouma, G., Krämer, and Zwarts, J., editors, *Cognitive Foundations of Communication*, pages 111–126.

Lassiter, D. and Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. *Semantics and Linguistic Theory*, 23:587.

Lassiter, D. and Goodman, N. D. (2015). Adjectival vagueness in a bayesian model of interpretation. *Synthese*, 194(10):3801–3836.

Qing, C. and Franke, M. (2015). Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. *Semantics and Linguistic Theory*, 24:23.

Sauerland, U. and Stateva, P. (2007). Scalar vs. epistemic vagueness: Evidence from approximators. *Semantics and Linguistic Theory*, 17:228.

Solt, S. (2015). An alternative theory of imprecision. *Semantics and Linguistic Theory*, 24:514.

Williamson, T. (1997). Précis of vagueness. *Philosophy and Phenomenological Research*, 57(4):921.

Égré, P. (2022). *Around "Around"*, page 75–86. Springer International Publishing.

Égré, P., Ripley, D., and Verheyen, S. (2019). The sorites paradox in psychology. In Oms, S. and Zardini, E., editors, *The Sorites Paradox*. Cambridge University Press.

Égré, P., Spector, B., Mortier, A., and Verheyen, S. (2023). On the optimality of vagueness: "around", "between" and the Gricean maxims. *Linguistics and Philosophy*, 46(5):1075–1130.