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

Adèle Hénot-Mortier (MIT), Steven Verheyen (EUR), Paul Égré & Benjamin Spector (CNRS, ENS-PSL, EHESS)

Research question

Numerical approximation expressions, such as ***around n*** and ***between x and y***, convey some uncertainty about an exact numerical value.

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

Yet, **(1a) is semantically vague while (1b) is not** (Égré, 2022; Égré et al., 2022). Égré et al., 2022 argue that on top of this, **(1a) and (1b) give rise to posterior distributions with very different shapes**.

In an online probability-elicitation experiment, we test a central prediction of the Bayesian account of ***around n*** vs. ***between x and x*** proposed by Égré et al., 2022. In this model, the probabilistic update produced by an imprecise ***around n***-sentence yields a posterior which is more "peaked" on the target value n than the posterior induced by a precise ***between n-a and n+a*** sentence.

Bayesian Model of *around* and *between*

Background. ***Around n*** and ***between x and y*** can be seen as determiners (cf. Eq. II and I). Eq. I is **relativized to a free parameter i, the half-length of the intended interval**.

$$[\![\textit{around } n]\!](P)(Q) = 1 \iff |\iota_{\text{MAX}}\, X.\, P(X) \wedge Q(X)| \in [n \pm i] \quad \text{(I)}$$

$$[\![\textit{between } x \textit{ and } y]\!](P)(Q) = 1 \iff |\iota_{\text{MAX}}\, X.\, P(X) \wedge Q(X)| \in [x; y] \quad \text{(II)}$$

Framework. A Bayesian listener $\mathcal{L}$ processing ***around n*** draws inferences about i and k, starting with a joint prior $\mathcal{P}_{\mathcal{L}}$ over (k, i) (cf. Lassiter and Goodman, 2013; Qing and Franke, 2015; Bergen et al., 2016). k and i are taken to be probabilistically independent.

***Around*-update.** Upon hearing (1a), $\mathcal{L}$ learns that $k \in [20 \pm i]$, and conditionalizes their joint distribution on this information. This yields. Eq. III (after some calculations).

$$\mathcal{P}_{\mathcal{L}}[k \mid \textit{around } n] \propto \mathcal{P}_{\mathcal{L}}[k] \sum_{i=|n-k|}^{n} \mathcal{P}_{\mathcal{L}}[i] \quad \text{(III)}$$

Under uncertainty about both i and k, $\mathcal{L}$ will reason that the sentence is more likely to be true if k is closer to 20. More generally, **numbers closer to n will receive a higher posterior probability than their prior probabilities**; while the opposite will hold for numbers further from n.

***Between*-update.** Upon hearing (1b), $\mathcal{L}$ conditionalizes their prior using the fact that $15 \leq k \leq 25$ (cf. Eq. IV).

$$\mathcal{P}_{\mathcal{L}}[k \mid \textit{between x and y}] \propto \begin{cases} \mathcal{P}_{\mathcal{L}}[k] & \text{if } k \in [x; y] \\ 0 & \text{if } k \notin [x; y] \end{cases} \quad \text{(IV)}$$

Key prediction

Given k_1, k_2, s.t. $|n - k_1| < |n - k_2|$, Égré et al., 2022 predict that for any prior, the ratio of the posteriors of k_1 and k_2 should be greater after an ***around n***-update than after an "equivalent" ***between x and y***-update (s.t. $[x; y]$ is centered on n and contains k_1 and k_2).

$$\underbrace{\frac{\mathcal{P}_{\mathcal{L}}[k_1 \mid \textit{around n}]}{\mathcal{P}_{\mathcal{L}}[k_2 \mid \textit{around n}]}}_{r_{a,k_1,k_2}} \geq \frac{\mathcal{P}_{\mathcal{L}}[k_1]}{\mathcal{P}_{\mathcal{L}}[k_2]} = \underbrace{\frac{\mathcal{P}_{\mathcal{L}}[k_1 \mid \textit{between x and y}]}{\mathcal{P}_{\mathcal{L}}[k_2 \mid \textit{between x and y}]}}_{r_{b,k_1,k_2}} \quad \text{(V)}$$

Focusing on ***around*** and ***between*** posteriors sharing the same support S, Eq. V can be averaged for all pairs $(k_1, k_2) \in S^2$ s.t. $|n - k_1| < |n - k_2|$ (we call this set of pairs K), yielding Eq. VI. **Graphically, Eq. VI implies that the *around* posterior should be more peaked than the *between* posterior.**

$$\underbrace{\frac{1}{|K|} \sum_{\substack{(k,k') \in S \times S \\ |n-k| < |n-k'|}} r_{a,k,k'}}_{R_a} \geq \underbrace{\frac{1}{|K|} \sum_{\substack{(k,k') \in S \times S \\ |n-k| < |n-k'|}} r_{b,k,k'}}_{R_b} \quad \text{(VI)}$$

Experiment

Goal. To test Eq. VI, we conducted an online experiment on Amazon MTurk. For each participant, we elicited and compared the posteriors generated by sentences of the form {***Around n*** / ***Between x and y*** } *people came to the party* (n=40, 50 or 60, randomized across participants).

Interval task.[1] The participant first defined the support of the posterior distribution induced by the two sentences (cf. Fig. 1 + Screens 1 & 2 in Tab. 1). To ensure that the ***around*** and ***between*** expressions were associated to similar supports, the bounds returned for the ***around n*** sentence were used as the *x* and *y* of the ***between x and y*** sentence.

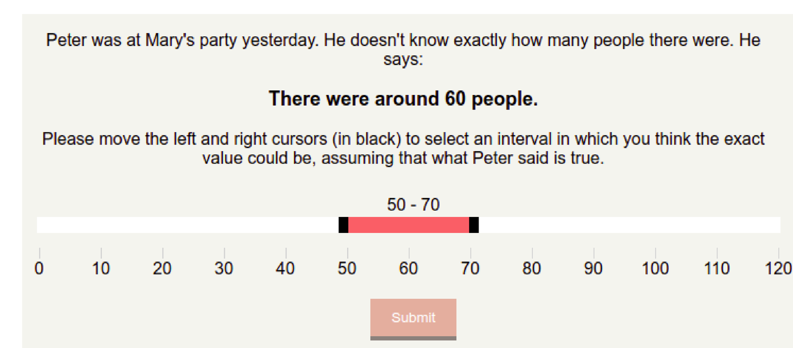

Figure 1. INTERVAL task

Histogram task. The participant then assigned relative weights to each value in the support defined during the related INTERVAL task (cf. Figure 2 + Screens 3/4 in Tab. 1). Weights were intended to provide empirical posteriors.

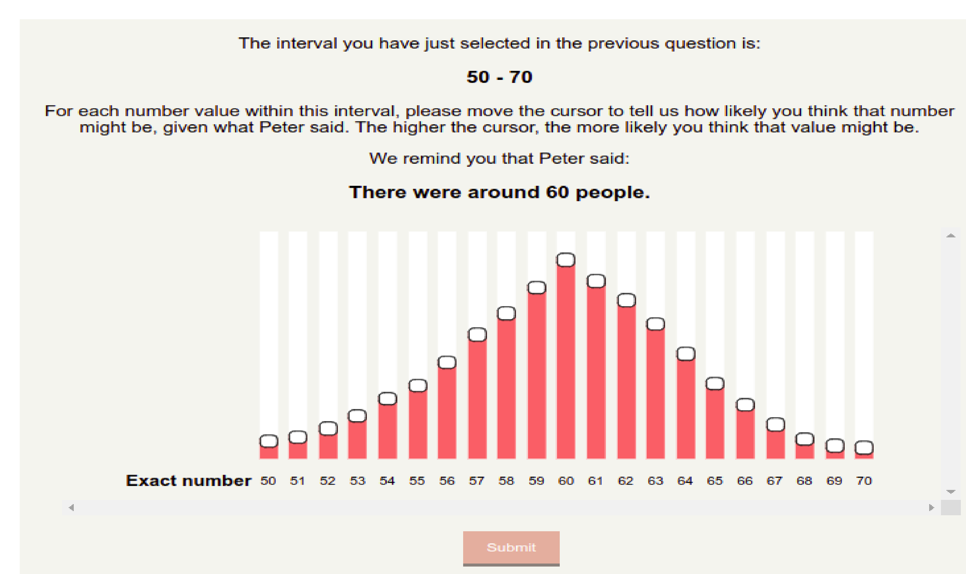

Figure 2. HISTOGRAM task

Screen	Expression	Task	Expected output
1	***Around 50***	INTERVAL	$[a; b]$ with $a \leq 50 \leq b$
2	***Between a and b***	INTERVAL	$[c; d]$ with $c = a$, $d = b$
3/4	***Around 50***	HISTOGRAM	$\mathcal{P}_{\mathcal{L}}[k \mid$ ***around 50***$]$
4/3	***Between a and b***	HISTOGRAM	$\mathcal{P}_{\mathcal{L}}[k \mid$ ***between a and b***$]$

Table 1. Arrangement of the tasks (n randomly set to 50)

Analysis. The hypothesis $R_a > R_b$, was verified for 162 participants vs. 78 ($p = 3.12 \times 10^{-8} \leq \alpha = .05$, one-tailed Sign test). The effect was of medium size (Cohen's $d = .24$). Post-hoc Bonferroni-corrected Wilcoxon tests run separately for each n were also significant at the $.05$ level.

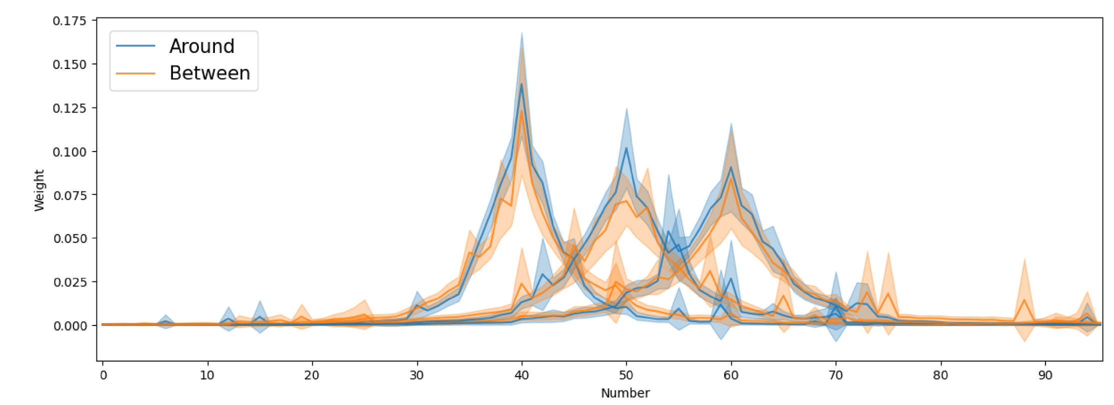

Figure 3. Mean posteriors + 95%-confidence envelopes (n=40, 50, 60)

Conclusion

Under the Bayesian model by Égré et al., 2022, the interpretation of ***around n*** in a given context is inherently probabilistic. **Thus, a vague *around n*-statement can communicate fine-grained probabilistic information in a way that its precise counterparts cannot.** Our study confirmed this prediction and constitutes the first empirical validation of the model. Future work may involve testing more targeted predictions of the model, by e.g. eliciting both priors and posteriors to check if the empirical updates match the predicted ones. Another avenue of work could consist in testing production rather than comprehension, by trying to validate our model not just for a literal listener ($\mathcal{L} = \mathcal{L}_0$) as we did here, but also for a higher-order speaker $\mathcal{S}_1$, who would be aware of the behavior of $\mathcal{L}_0$.

Selected references

Channell, J. (1994). *Vague language*. Oxford University Press. Williamson, T. (1997). Precis of vagueness. *Philosophy and Phenomenological Research*. Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *L&P*. Frazee, J., & Beaver, D. (2010). Vagueness is rational under uncertainty. In *Logic, language and meaning*. Cummins, C., Sauerland, U., & Solt, S. (2012). Granularity and scalar implicature in numerical expressions. *L&P*. Lassiter, D., & Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. *SALT*. Qing, C., & Franke, M. (2015). Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. *SALT*. Solt, S. (2015). An alternative theory of imprecision. *SALT*. Bergen, L., Levy, R., & Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*. Égré, P. (2022). Around "around". In *Measurements, numerals and scales – essays in honor of stephanie solt*. Égré, P., Spector, B., Mortier, A., & Verheyen, S. (2022). On the optimality of vagueness: "around", "between", and the gricean maxims. *L&P*.

[1] We thank Athulya Aravind for suggesting to split the tasks in that way, which allowed to randomize the HISTOGRAM tasks and ensured that order effects were kept minimal.

mortier@mit.edu