GLOSSA psycholinguistics

Katzir, N., & Ariel, M. (2026). Don't overthink! Round number expressions are interpreted sharp. *Glossa Psycholinguistics*, 5(1): 8, pp. 1–49. DOI: https://doi.org/10.5070/G6011.48658

eScholarship
University of California

Don't overthink! Round number expressions are interpreted sharp

Nicole Katzir, Centre for Language Studies, Radboud University; Linguistics Department, Tel Aviv University, nicole.katzir@ru.nl

Mira Ariel, Linguistics Department, Tel Aviv University, mariel@tauex.tau.ac.il

The interpretation of round numerals inspired two competing hypotheses. The first, the preferred-approximation hypothesis, claims that round numerals are more likely to be interpreted as approximate, while sharp numerals are interpreted as precise by default (Krifka, 2002, 2007). This hypothesis is based on the idea that addressees recognize that speakers can use round numerals for approximation more readily than sharp ones, and, therefore, derive the approximate interpretation through rational inferencing. The second hypothesis, the precise-by-default hypothesis, holds that all numerals, round and sharp alike, are interpreted as precise by default (Ariel, 2021; Ariel & Levshina, 2024). The present study investigates the relationship between round counting numerals and approximation through four experiments that directly compare round and sharp numerals across different experimental settings. Our findings show that participants do recognize speakers' license to use round numerals imprecisely more than sharp ones. However, this preference does not lead them to actually interpret round numerals imprecisely. We offer several explanations for this gap.

1. Introduction

Round number expressions tend to be simpler in form than sharp ones, and corpus findings show that they are also more frequent than sharp numerals of the same magnitude (Coupland, 2011; Dehaene & Mehler, 1992; Dorogovtsev et al., 2006; Gega et al., 2021; Jansen & Pollmann, 2001; Woodin et al., 2023).[1] The verbal simplicity and high frequency of round numerals go hand in hand with a cognitive advantage: round numbers stand for more culturally salient concepts, which is why what counts as a round number varies cross-linguistically. The higher frequency of round numerals makes them easier to access and to process (Balota & Chumbley, 1984; Gega et al., 2021; Mason et al., 1996; Oldfield & Wingfield, 1965; Solt et al., 2017). But do these differences also correlate with interpretative differences? Specifically, are round numerals interpreted as approximate numerosities more often than sharp numerals? Scientists and laypersons alike share an intuition that round numerals are associated with approximate interpretations.[2] Krifka (2002, 2007) and others (e.g., Lasersohn, 1991) have argued that the preferred interpretation of round numerals is approximate, a stance that has been accepted by many (Bastiaanse, 2011; Woodin et al., 2023, to name a few). But Ferson et al. (2015), Ariel (2017, 2019, 2021; Ariel & Levshina, 2024), and Gega et al. (2021) have argued that *all* numerals are interpreted as precise by default. The goal of this article is to directly compare the interpretations of round and sharp numerals in order to decide this issue.

1.1 Theoretical background

The two theoretical positions differ primarily in their predictions about the interpretation of round numerals. Mathematically, roundness is defined by a set of structural properties tied to the base of the number system used in the relevant culture (Sigurd, 1988). The mathematical features contributing to roundness for counting systems using base 10 are simple multiples of the base number and simple fractions of the base number (Lotz, 1955; Sigurd, 1988). These include whether a numeral can be expressed as a single-digit multiple of a power of 10 (*10-ness*), five times a power of 10 (*5-ness*), two times a power of 10 (*2-ness*), or 2.5 times a power of 10 (*2.5-ness*). These properties do not correspond to ordinary divisibility. For example, *40* has 10-ness, 5-ness, and 2-ness; *50* has 10-ness and 5-ness; *38* has none. The number and type of roundness features predict a numeral's frequency in use (Woodin et al., 2023), suggesting that these mathematical properties also have cognitive consequences. In domains with different underlying bases, e.g., time, other values count as round accordingly.

[1] We distinguish between *numerals*, or number words, and numbers. Numerals are the linguistic expressions that encode the mathematical number concepts.

[2] For example, according to Wu and Aparicio (2025, p. 435), "imprecision is pervasive in everyday communication, manifesting across a wide range of lexical items, such as maximum standard absolute adjectives…, or *round numerals* (e.g., one hundred), among others" (emphasis added).

With this notion of roundness in place, we turn to the rationale behind the preferred-approximation and the precise-by-default positions. Starting with the preferred-approximation hypothesis, one might think that round numerals tend to convey approximate values because they are much more frequent in discourse than could possibly be justified by actual states of affairs (see the findings in Krifka, 2007, regarding the relative frequency of the rounder 50 versus the less round 40 and 60 in Norwegian versus Danish, and Basque). Speakers obviously actively choose to round numbers, which means that many round numerals in discourse do not correspond to factually true round numbers. But what about the addressee's interpretation?

The preferred-approximation proposal is based on Bayesian inferences and game-theoretic mechanisms (Bastiaanse, 2011), since, according to this position, addressees infer the intended meaning by engaging with the speakers' point of view, their motivations, and the alternatives available to them. According to Krifka (2002, 2007), these alternatives are a function of an underlying scale and its degree of coarseness. The degree of coarseness is determined by the minimal units of the scale. Thus, a scale such as [39–40–41–42] is more fine-grained than a scale such as [35–40–45–50], because the distance between the units is smaller. Hence, each unit on the fine-grained scale stands for a smaller scalar range. When the scale that determines the degree of precision of the sentence is fine-grained, a numeral like 40 will receive a precise interpretation. Under the coarse-grained scale, however, it will be imprecise. On this account, round numerals are ambiguous between precise and imprecise interpretations. However, imprecise utterances of round numerals are generally favored, because round numerals are simpler. According to Krifka, a speaker wishing to convey an interval, such as [39–40–41–42], has four alternative numerals to choose from (on the assumption that a speaker may actually convey such an interval by uttering one of its members). However, only *forty* has an advantage over the other alternatives, having the simplest form. This is recognized by the addressee decoding the speaker's utterance. Additionally, imprecise utterances offer more values, which increases the likelihood of the statement's being true: an exact value has a lower probability of being true than an interval. Furthermore, Krifka (2002) suggests that a more coarse-grained representation might be less cognitively costly. Taken together, these assumptions lead to the conclusion that, for round numerals, the imprecise, approximate interpretation is more likely. This is the first theoretical hypothesis this article aims to test.

Complementing Krifka's motivations for why speakers may choose to use round numerals for imprecise meanings is Van Der Henst et al.'s (2002) Relevance-Theoretic argument that rounding is a hearer-oriented strategy. According to Relevance Theory (Sperber & Wilson, 1986), speakers opt for optimally relevant utterances, i.e., ones that allow the listener to derive a sufficient amount of cognitive effects using only a minimal processing effort. For example, when someone is asked to tell the time, if they are being helpful, they should provide an answer that will be useful for the asker (to the best of their knowledge), while imposing the least amount of cognitive load on the asker. This should lead them to prefer a round time. Thus, using approximative values in contexts where precision is not crucial reduces the cognitive effort imposed on the addressees, while broadly maintaining the contextual implications derived from the utterance. While Van Der Henst and colleagues do not explicitly deal with the addressee's interpretation, it's possible to apply the model proposed by Krifka here, as well, by positing that addressees may also consider the speaker's motivation for an optimally relevant utterance when choosing a numeral.

Moving on to the precise-by-default camp, the starting point for this position is the plausibility of the assumption that there must be a functional motivation behind the extremely rich number system culturally evolved in most languages. Presumably, it complements our species-inherited cognitive propensity to only assess quantities approximately (Feigenson et al., 2004; Odic & Starr, 2018). The number line, based on the successor (+1) principle, makes it possible for speakers to provide precise descriptions regarding numerosities. Since sharp and round numerals figure equally on the number line, both should prompt precise number interpretations by default.

Ariel (2017, 2019, 2021; Ariel & Levshina, 2024) has argued that typical lexemes and numerals belong to different lexical domains, such that the former more easily give rise to contextually-adjusted (nonliteral) interpretations (Carston, 2002). The meaning of *nice*, for example, can be broadened to mean 'kind of nice, close to attractive/kind', etc. According to this account (see especially Ariel & Levshina, 2024), typical lexemes belong to sparse semantic domains, where the relevant meaning options are not continuous over the domain (i.e., the member meanings do not fully cover the whole meaning space). Instead, there are no-man's-land meaning spaces between adjacent competing members, which broadening can take advantage of.

Unlike the sparse semantic domains characteristic of non-numeral lexemes, the counting number line constitutes a dense lexical domain, according to Ariel. Dense semantic domains are continuous and show no no-man's-land meaning spaces between adjacent members.[3] Therefore, the interpretation of numerals is strict and defaults to precision. This is the second theoretical hypothesis we wish to examine. Contra Bastiaanse (2011), Ariel's reasoning was that in order to trigger an approximate interpretation, numerals must be explicitly marked (by dedicated approximators, such as *about, around,* etc.) (see also Gega et al., 2021). In this case, it is actually the modifier, rather than the numeral, that is (compositionally) responsible for the created range, resulting in an approximate reading.

According to this view, it is not impossible, however, to construct a sparse system for numerals, such as 10–20–30 (moving to a coarser scale granularity, according to Krifka, 2007). The spaces between, say, 10 and 20, and between 20 and 30 can then be designated as no-man's-land

[3] Note that the density relevant here is *not* the assumption that there's an infinite number of *numbers* between any two (non-integer) numbers.

meaning spaces. In this way, *20*, for example, can broaden towards *10* and towards *30* (say, '16–25'). However, according to Ariel & Levshina (2024), interlocutors avoid such a domain shift. They assume that by default, it is the dense, +1 numeral domain that is relevant for interpreting both round and sharp numerals, where the minimally different members (+1, –1) block meaning broadening. Creating a pragmatic halo (Lasersohn, 1991) is not a common step then (especially for *counting* numerals).[4] This is contrary to Krifka's position that coarse scales are less cognitively costly and, therefore, preferred for round numerals.

1.2 Previous findings

This subsection reviews the empirical evidence that supports the two positions above, starting, again, with the preferred-approximation position. One of the earliest studies on numerical use for approximation was Van Der Henst et al. (2002), who found that people were much more likely to provide a round answer (e.g., 6:15 when the actual time was 6:13) when asked for the time. Similar results were obtained by Mühlenbernd & Solt (2022), who found that participants round off the time of a car accident when they know the precise time it happened, and they do that more often in the context of a conversation with a neighbor than in a witness statement in a police station. Using a different methodology but with similar conclusions, Woodin et al. (in prep.) examined the numbers printed on jigsaw puzzle boxes as indicating the number of pieces. They found that these numbers are preferably round and often somewhat imprecise. These studies provide support for the claim that speakers tend to use round numerals for approximation.

The idea that round numerals are easier to process, which serves as a motivation for speakers to use them for approximate values, finds support in Mason et al. (1996), which showed that participants performed better in recall and recognition tasks involving round versus sharp numerals. Similarly, Solt et al. (2017) found that participants who were shown a sequence of clock times were able to determine whether a certain time appeared as part of the sequence with higher speed and accuracy for round times (3:15, 8:30) than for sharp ones (3:21, 8:36). These results indicate that round values have an advantage in short-term memory over non-round values. In an additional time-arithmetic task, participants were more accurate and responded more quickly when the starting time was round than when it was non-round, as well as when the temporal increment was round versus non-round, providing further support for the claim that round numbers are more readily processed.

[4] It has been argued that numerals in measurement phrases are less precise than those in count phrases. Cummins (2015, Chapter 6) notes that measurement scales are divisible into many sub-scales, and, therefore, rounding is performed to the nearest scale unit, instead of multiples of 5 or 10. Ferson et al. (2015) and Ariel and Levshina attribute their interpretation as 'approximate' to assumed measurement errors.

Kao et al. (2014) present some evidence for the addressee's recognition of the speaker's tendency to round numbers off. The study used prices of items to test non-literal uses of numerals (both approximation and hyperbole). Participants read scenarios in which speakers purchased an item and were asked by a friend, "Was it expensive?". They then replied that the item had cost *N,* where *N* was either a round or a sharp number. The participants were then asked to rate the likelihood of the quoted price representing an exact, approximate, or hyperbolic value. The results showed that the average bias for exact interpretation is significantly higher for sharp than for round numerals, indicating that participants recognize that when speakers use round numerals, they are more likely to diverge from the state of affairs than when they use a sharp numeral.

Moving to the evidence in favor of the precise-by-default position, a corpus study by Ariel (2021) showed that whereas typical lexemes (*nice, blue, green, red*) were hardly ever overtly broadened in her corpus (~3%), a select set of numerals (*2, 7, 10, 17, 30*) were explicitly broadened 8 times as often (~25% of the time, based on the Santa Barbara Corpus (Du Bois et al., 2000) and the Longman Corpus of Spoken American English. Comparing the co-occurrence of one Hebrew approximator (*be.erex* 'approximately') with numerals and with a maximum standard absolute adjective (*rek*, 'empty'), Ariel and Levshina (2024) showed a dramatic difference: there were virtually no 'approximately' modifiers among the modifiers of Hebrew 'empty' (1.1%). But virtually all of the numeral modifiers involved 'approximately' (94.4%). Assuming that default interpretations need not be marked, whereas non-default ones might need to be explicitly marked, these findings suggest that numerals are interpreted as precise numerosities by default. Note also that approximators, such as *about*, are far more frequent than "precisifiers" (such as *exactly*) as numeral modifiers.[5] Furthermore, Jansen and Pollmann (2001) found that round numerals are modified by *about* and its Dutch (*ongeveer*), French (*environ*), and German (*etwa*) counterparts much more frequently than a chance distribution would suggest.

Aparicio's (2017) series of self-paced reading studies (Experiments 3a, 3b, 4, 6a, and 6b), which did not find increased reading times for sentences with round numerals in precise contexts, failed to find support for imprecise interpretations of round numerals. These contexts included cases where a question preceding the sentence with the numeral specifically called for either a precise or an approximate interpretation (Experiment 3a), when the numeral was modified by *exactly/about/approximately* (Experiments 3b, 4), and when the general context was biased for either a precise or an approximate interpretation (Experiments 6a and 6b). The

[5] See Gega et al. (2021) and Ariel and Levshina (2024) for corpus counts demonstrating the rarity of *exactly* versus the frequency of approximators as numeral modifiers. For example, the BNC contains 23 tokens of *exactly three/3*, which constitute 0.023% of all *three/3* tokens (100,279), but the 13,888 *about three/3* tokens constitute 1.4% of *three/3* (600 times more). The same search for *fifty/50* yielded a smaller, but still highly dramatic, gap in favor of *about* versus *exactly* numeral modifiers (97 times more).

lack of a reading time penalty indicates that there is no evidence that processing approximate interpretations of round numerals is facilitated compared to precise interpretations. In fact, the significant differences that were observed in the experiment showed that precise interpretations were processed faster.

Note that Ferson et al. (2015) also found that bare numerals are overwhelmingly interpreted as precise, but their bare numeral targets included both round and sharp numerals, count and measurement numeral phrases, and integer and decimal number expressions. These different numeral types may very well bias towards more or less precise readings, but Ferson and colleagues do not present separate counts for the different numeral types, so the overall result may not be informative enough. Similarly, Ariel and Levshina (2024) argued for a precise interpretation for all numerals, where again, no systematic attempt was made to distinguish round and sharp numerals. However, a study by Solt and Stevens (2018), in which participants read sentences with numerals modified by *about*, *some* or without modification, and were asked to specify the range of numbers conveyed by the sentence, did compare round and sharp numerals directly. The results showed that 80% of the responses to unmodified numerals counted as exact, for both sharp and round numerals. However, due to the lack of a detailed description of the study's materials, the results are difficult to assess. Hence, the experiments reported below were designed to directly compare round and sharp count numerals of similar magnitudes.

This line of thinking receives additional support from experiments probing the interpretation of maximum standard absolute adjectives, such as *empty* or *straight*. Although they correspond to an end point of a scale, they can shift from a boundedness to a graded schematicity even without special marking (Paradis, 2001), and can be more or less precise (Kennedy, 2007). Both Ronderos et al. (2024) and Wu and Aparicio (2025) have argued that, in the absence of contextual prompts for approximation, such adjectives were primarily interpreted as precise. Importantly, interpreting deviating (approximation-inducing) cases incurred a processing cost (see also Aparicio, 2017), which provides another piece of evidence for the assumption that maximum absolute adjectives are interpreted as precise by default. Still, some deviations from the literal denotation were accepted by Ronderos and colleagues' participants (between 18%–89% in their Experiment 1). Moreover, Wu and Aparicio found participants' acceptance for both interpretative strategies (at ∼20%) when participants' attention was drawn to an interpretative disagreement between two speakers regarding precise versus approximate interpretations. Additionally, Ronderos et al. (2024) found that a small deviation from a precise interpretation of maximum standard absolute adjectives was acceptable to participants at a rather high rate (68%), even when the experimental context explicitly biased participants to opt for *strict* interpretations. The picture that emerges from the interpretation of maximum standard absolute adjectives shows a clear bias for preciseness, although approximations are not totally ruled out. According to the precise-by-default position, the interpretation of numerals should be even more strict, since they belong to a dense lexical domain.

All in all, different empirical findings lend support to various components of the two competing proposals. Nevertheless, a systematic experimental investigation that directly tests the predictions of the two competing theoretical accounts has not been conducted, as far as we know. The next subsection presents the current study, which aims to do that.

1.3 The present study

In 1.1, we introduced two theoretical proposals regarding the interpretation of round numerals. The rationales behind both competing proposals seem to offer reasonable guiding principles for how to pragmatically adjust (or not adjust) the baseline precise interpretation of numerals. According to the preferred-approximation position, addressees reason about how states of affairs must be, as well as about the speaker's motivations and available alternative utterances. Since round numerals are simpler than sharp ones, addressees expect speakers to use them when the precise value is not known or not crucial. This leads them to adopt approximate interpretations for round numerals. On the other hand, according to the precise-by-default position, numerals belong to a dense lexical field, and, therefore, broadening (or approximating) is not common. Thus, addressees are expected to not adjust the precise numeral meaning, which is linguistically derived. The goal of this article is to find out which of these competing proposals better predicts how addressees actually interpret round numerals.

Given the two competing proposals, we derive the following predictions:

(i) If addressees follow the preferred-approximation reasoning, round and sharp numerals should elicit different responses. Round, but not sharp, numerals would be predominantly interpreted as approximate.

(ii) If addressees follow the precise-by-default account, based on the fact that all numerals are members of the dense number line, there would be no difference between round and sharp numerals, both receiving precise interpretations predominantly.

(iii) Addressees might oscillate between the two strategies, in which case a mixed interpretative pattern would emerge (either due to an inconsistent adoption of the two strategies by all or most addressees, or due to consistently different preferences for one or the other strategy by individual addressees).

We recognize that precise and approximate readings are not the only readings potentially associated with numerals. All number expressions receive a variety of context-relevant interpretations: 'exactly N', 'at most N', 'at least N', and 'approximately N'. Interpreting numerals, addressees start with their linguistic meaning and then adjust it to derive the speaker-intended reading in the specific context. Most linguists assume a lower-bound only lexical meaning for numerals ('at least N'), which is then strengthened by an upper-bounding ('no more than N') implicature. The result is an 'exactly N' representation. Originally, these implicatures were considered pragmatic, although the relevant implicature was classified as a generalized conversational implicature, namely, one that goes through unless canceled by context (Horn, 1989; Levinson, 2000). A variant account views the relevant implicature as a grammatical implicature (Chierchia et al., 2012). A third account takes the numerals' lexical meaning to be lower- and upper-bounded, so no upper-bounding implicature is necessary (Ariel, 2006, and onwards; Breheny, 2008). The differences between these theories are orthogonal to our research question, because all three agree that, prior to the consideration of context-specific inferential processes, numerals are taken to have a precise interpretation. We here ignore 'at least' and 'at most' readings, and focus on a potential difference between round and sharp numerals with respect to approximate and precise interpretations.

We present three main experiments below. Experiment 1 tests the assumption of the preferred-approximation hypothesis that interlocutors are aware of speakers' license to diverge from precise numerical descriptions, favoring an imprecise use of round numerals, in particular. Our results support this assumption. Still, the question remains whether addressees actually proceed to then engage in the reasoning outlined by the preferred-approximation account, which should lead them to adopt approximate interpretations.

Experiments 2a and 2b aim to find out whether listeners accept an intended interpretation of a range of values for round numerals more than for sharp ones by comparing bare (unmodified) round and sharp numerals with numerals modified by *about* (necessarily approximate) and by *exactly* (necessarily precise).

While Experiments 2a and 2b test the interpretation of numerals, Experiment 3 utilized a truth-value judgment (TVJ) task to test the compatibility of numerals with approximate interpretations. In this task, participants are presented with a sentence containing a numeral, followed by the relevant state of affairs. Their task is to judge whether the sentence is true or false in light of the state of affairs. TVJ is a more "tolerant" diagnostic for interpretation, because it may be enough for the state of affairs to be compatible with (rather than identical to) the statement. If round numerals are more compatible with approximate interpretations than sharp ones, we should see higher acceptance rates for utterances with round numerals in light of a state of affairs in which the actual numeral is not the same one as in the original utterance, but is within the numeral's range of compatibility.

Based on the results of Experiments 2a, 2b, and 3, we will argue that listeners actually do not translate their awareness of speakers' rounding practices into an interpretative strategy of approximation. Just like sharp numerals, round ones too are overwhelmingly interpreted as 'precise N'. In order to reconcile these seemingly contradictory results, we propose that when interlocutors are consciously asked about imprecise numeral descriptions (Experiment 1), they show awareness of the advantage of using round numerals. But left to their own devices, they don't "overthink". They don't engage in reverse engineering, so their default interpretation of both numeral types is as precise.

Decontextualized tasks were used in all the experiments below. Specifically, participants were not offered any cues for why the number information was relevant. Such cues might very well affect participants' tolerance for imprecision. Participants were also not provided with cues about the speaker's personality, e.g., as pedantic versus laid-back, which has been identified as relevant for acceptance rates of round numerals diverging from the factual number (Beltrama & Schwarz, 2022, and the references therein). In other words, the design attempted, as much as possible, to strip away the subjective aspects of numeral use and its crucial contribution to argumentation (in the sense of Anscombre & Ducrot, 1989). The only interest is in the most prominent interpretation of round numerals, for which there are competing claims (all agree that round numerals *can* be interpreted both precisely and approximately). These conflicting predictions have not been tested adequately, i.e., by directly comparing round and sharp numerals across different experimental tasks that tap into interpretation.

2. Experiment 1

The aim of Experiment 1 is to test the assumption that interlocutors are aware of the speakers' license to use round numerals imprecisely, i.e., to diverge (to some extent) from the actual state of affairs. Such awareness constitutes a necessary condition for inferring an approximate interpretation for round numerals. To test this, we used a task in which participants were presented with a sentence containing a round or a sharp numeral, followed by a fact about the actual numerosity. Participants were asked to judge whether the speaker's utterance was a reasonable description of the fact. If round numerals are more easily seen as compatible with range values (pointing to a possible approximate interpretation) than sharp ones are, we should see higher acceptance rates for utterances with round numerals where the numeral in the fact slightly diverges from the numeral in the utterance, as compared with utterances containing sharp numerals.

2.1 Methods

2.1.1 Participants

The participants for all experiments in this study were recruited on Prolific (prolific.com) and gave their consent to participate in the study. 420 native speakers of American English (205 female, mean age = 30.85, SD = 5.61) participated in Experiment 1.

2.1.2 Materials

The materials consisted of stimulus sentences and facts. The experiment followed a 2 × 4 design, with numeral type (round/sharp) manipulated between subjects and condition (matching/

within range/out of range/false) manipulated within subjects. The four conditions were defined by the numerical distance between the sentence and the fact numerals.

Numeral type (round/sharp) was manipulated between subjects. Our primary motivation for this manipulation between participants was to reduce the salience of the contrast between round and sharp numerals, which we worried might otherwise invite strong extra-linguistic intuitions about imprecision.

For sentences containing round numerals, we used *20*, *30*, *40*, *60*, *70*, *80*, and *90*, and for sharp numerals, we used the same numerals plus 2 (*22*, *32*, etc.). Each numeral appeared only once in each list and appeared in all four conditions across the experimental lists. Each participant was exposed to one of 14 experimental lists, each containing 7 experimental items. No filler items were used. The order of items was randomized per participant. A full experimental set for round numerals can be found in **Table 1**. A full experimental set for round numerals can be found in the supplementary files at the OSF site for this article.

Table 1: Materials used in the round numeral condition of Experiment 1.

Sentence	Fact	Condition
Mary: There are 20 history books on the shelf	There are 20 history books on the shelf	Matching
John: There are 30 marbles in the red bowl	There are 33 marbles in the red bowl	Within range
Michael: There are 40 cars in the actor's garage	There are 37 cars in the actor's garage	Within range
Dan: There are 60 pages in the police file about the car accident	There are 67 pages in the police file about the car accident	Out of range
Suzanne: There are 70 flowers in the Chinese vase	There are 63 flowers in the Chinese vase	Out of range
Amy: There are 80 paintings in the art gallery in Paris	There are 94 paintings in the art gallery in Paris	False
Christine: There are 90 TV sets in the store on Washington Street	There are 76 TV sets in the store on Washington Street	False

Four within-subject conditions were used, composed of pairs of sentences and facts. All sentences and facts followed the form "There are N OBJECTS in the LOCATION."[6] All objects were inanimate, based on the assumption that animate nouns might raise expectations for precision, as individual animate entities tend to be more prominent.

[6] One of the reviewers suggested that existential *there are…* constructions may favor lower-bounded interpretations of numerals more than is typical. Even if this is the case, both numeral types occur in the same construction in our materials, so this should not affect their comparison.

In the matching condition, the sentence and the fact contained the exact same numeral – i.e., there was no divergence. In a within range condition, the fact contained a numeral 3 places above or below the numeral in the sentence, e.g., 33 and 27 for *30* (two items total). Based on the assumption that, in a base-10 counting system, rounding often occurs up to the nearest 10, values below ± 5 are expected to fall within the range of compatibility of a round numeral. We took into account two further factors in choosing the specific six-digit interval of compatibility for our target numerals (e.g., 37–43 for the round *40* and 39–45 for the sharp *42*).[7] First, we wanted to make sure that not just the round numeral targets were even numbers, but also the sharp ones. This is because studies on numerical cognition indicate that even and odd numbers are processed differently (Nuerk et al., 2004). In addition, since the lower bound of the interval for the round numeral necessarily involved a lower decade (e.g., **2**7 for ***3**0*), we made sure that the same applies to the sharp number targets. To do so, we selected sharp numerals ending in the digit *2*, e.g., ***32***. This way, the lower bound of their interval would also cross the decade boundary (e.g., **2**9 for ***3**2)*. This also guaranteed a rather minimal magnitude difference between the sharp and round targets. A four- or a one- or a two-digit interval would fail to meet these conditions.

Next, in the out of range condition, the fact contained a numeral 7 places above or below the numeral in the sentence, e.g., 67 and 53 for *60* (two items total), which should be outside the range of compatibility. In the false condition, the fact contained a numeral 14 places above or below the numeral in the sentence, e.g., 94 and 66 for *80* (two items total).

2.1.3 Procedure

This experiment, as well as the other experiments reported here, was administered using PCIbex (Zehr & Schwarz, 2018).

Participants read a sentence uttered by a speaker, followed by a fact. They were then asked to make a binary choice about whether the speaker's choice of words was "a reasonable description" of the fact.

[7] One reviewer noted that in some non-matching conditions, the numeral appearing in the fact (e.g., 75) has 5 as its single digit and might, in principle, be interpreted imprecisely. However, such numerals lack the canonical roundness features discussed in Section 1 and do not show as high a frequency relative to other sharp numerals as multiples of 10 do. We therefore do not expect them to lead to frequent approximate interpretations under the preferred-approximation hypothesis. Moreover, this is the case for only a small subset of the trials, i.e., sharp numerals in the within-range condition, when the fact numeral is above the sentence numeral.

2.1.4 Predictions

The experiment tests whether addressees recognize the speakers' license to use round numerals imprecisely, more than sharp ones. If they do, round numerals should yield higher acceptance rates than sharp numerals in cases of a small divergence between the sentence and the fact. This difference should be most prominent in the within range condition: using round numerals (e.g., *30* for '27' or '33' states of affairs) is expected to be accepted more often than sharp numerals (e.g., *32* for '29' or '35' states of affairs). In the matching condition, both numeral types should be equally accepted, as the sentence and fact match exactly. In the false condition, where the divergence is substantial (±14), both should be equally rejected.

The out of range condition (±7) is more exploratory. If a 7-unit discrepancy is already too large for approximation to be felicitous, both numeral types should yield similarly low acceptance rates. However, if rounding is still considered plausible at this distance, we expect higher acceptance rates for round numerals. This outcome would suggest that addressees not only recognize round numerals as approximate, but also allow a relatively wide tolerance window for approximation.

2.1.5 Analyses

All analyses in this article were conducted in R (R Core Team, 2021). We fit a Bayesian mixed-effects logistic regression model using the brms package (Bürkner, 2021), with response as a binary variable (yes/no) as a function of numeral type (round/sharp), condition (matching/within range/out of range/false), and their interaction. Categorical predictors were sum-coded. Following Barr et al. (2013), the model included a maximal random-effects structure with random slopes and intercepts for participants and items. The model used the default weakly informative priors supplied by brms, namely normal(0, 1) priors for fixed effects, a Student-t(3, 0, 2.5) prior for the intercept, Student-t(3, 0, 2.5) priors for random-effects standard deviations, and an LKJ(1) prior for random-effects correlation matrices. Four MCMC chains were run for 4000 iterations each (1000 warmup). Convergence was assessed using the R-hat diagnostic and visual inspection of trace plots. We also inspected the magnitude of the fixed-effect estimates and their standard errors, to ensure that ceiling or floor effects did not lead to unstable estimates. Regression coefficients are reported as posterior β estimates (medians) on the log-odds scale, with 95% credible intervals.

Planned pairwise comparisons were carried out using emmeans (Lenth, 2022) applied to the posterior distributions of the fitted model. All contrasts are reported as posterior medians with 95% credible intervals.

2.2 Results

The results of Experiment 1 are plotted in **Figure 1**.

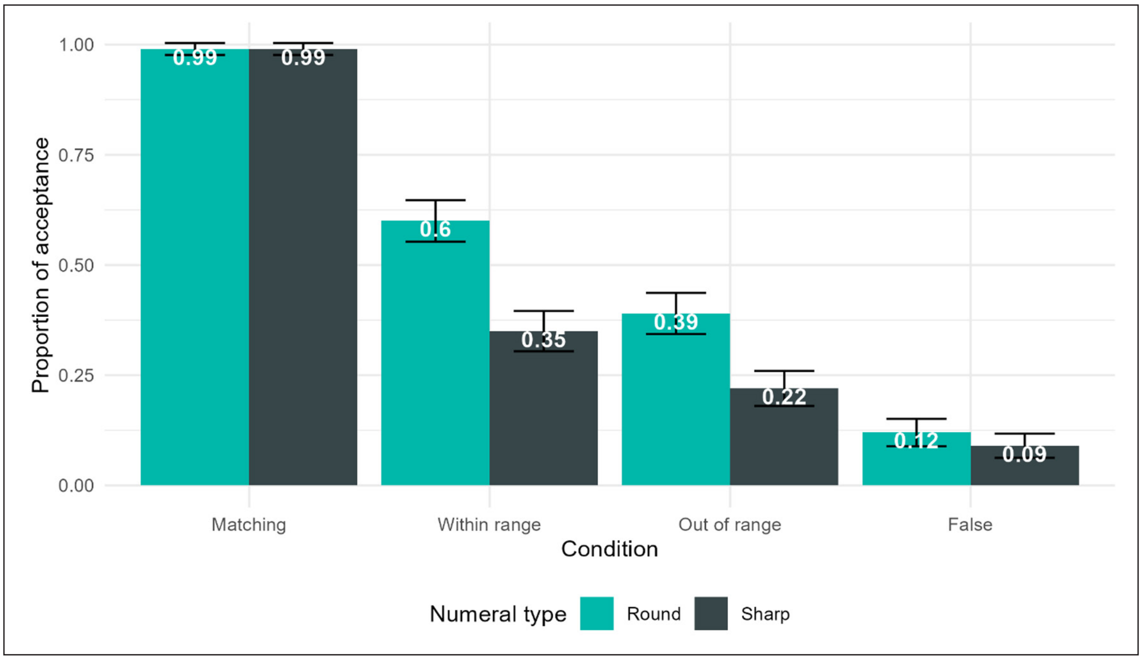

Figure 1: Results of Experiment 1. Error bars represent 95% CI.

The model coefficients are presented in **Table 2**, and the model's posterior estimates are visualized in **Figure 2**.

Table 2: Bayesian mixed-effects logistic model, formula: response ∼ type * condition + (1 + type * condition | participant) + (1 + type * condition | item). Factors are sum-coded. Therefore, the intercept reflects the grand mean.

Predictor	Estimate	Std. Error	95% CI	$\hat{R}$
Intercept (Grand Mean)	0.39	0.41	[–0.33, 1.30]	1.00
Type: Round	0.69	0.32	[0.06, 1.34]	1.00
Condition: False	–4.11	0.47	[–5.17, –3.32]	1.00
Condition: Matching	7.04	1.15	[5.24, 9.71]	1.00
Condition: Out of Range	–2.14	0.45	[–3.14, –1.41]	1.00
Round × False	–0.32	0.32	[–0.95, 0.33]	1.00
Round × Matching	–0.45	0.66	[–1.75, 0.90]	1.00
Round × Out of Range	0.10	0.28	[–0.46, 0.64]	1.00

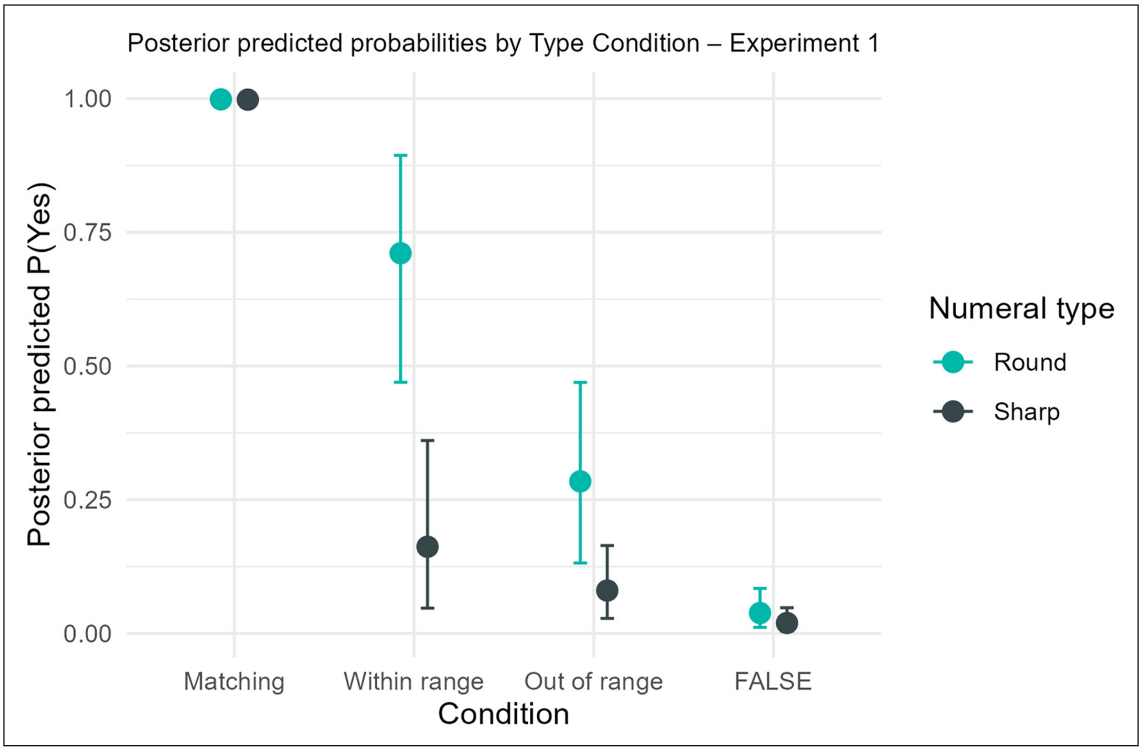

Figure 2: Posterior predicted probabilities of a 'Yes' response by numeral type and condition. Points represent the posterior mean estimates from the Bayesian mixed-effects model: response ~ type * condition + (1 + type * condition | participant) + (1 + type * condition | item). Error bars represent 95% Highest Posterior Density (HPD) credible intervals.

The model revealed that the acceptability of numeral type depended on condition (a credible interaction). While round and sharp numerals were accepted at equal rates in the False and Matching conditions, pairwise comparisons showed that round numerals were accepted significantly more than sharp numerals in the within-range (estimate = 2.48, *CI* = [1.18, 3.75]) and out-of-range (estimate = 1.81, *CI* = [0.33, 3.42]) conditions. Full comparisons are presented in **Table 3**.

Table 3: Pairwise comparisons of numeral type within each condition.

Condition	Contrast	Estimate	95% CI
False	Round – Sharp	0.89	[–0.37, 2.26]
Matching	Round – Sharp	–2.53	[–18.38, 9.37]
Out of range	Round – Sharp	1.81	[0.33, 3.42]
Within range	Round – Sharp	2.48	[1.18, 3.75]

Pairwise comparisons between conditions revealed credible differences in acceptability. The matching condition led to the highest acceptance rate of all conditions. False sentences were the least accepted, lower than within-range, out-of-range, and matching (all effects are credible). Within range sentences were more accepted than out of range, but less than matching. The full comparisons are presented in **Table 4**.

Table 4: Pairwise comparisons of condition in Experiment 1 (averaged across numeral type).

Contrast	Estimate	95% CI
False – Matching	–10.94	[–14.39, –8.54]
False – Out of range	–1.96	[–2.79, –1.15]
False – Within range	–3.30	[–4.38, –2.34]
Matching – Out of range	8.95	[6.39, 12.22]
Matching – Within range	7.61	[5.17, 11.14]
Out of range – Within range	–1.35	[–2.20, –0.49]

2.3 Discussion

The experiment tested whether addressees are sensitive to the speakers' license to use round numerals imprecisely, more so than sharp ones. The findings indicated that they are. This is supported by the fact that round numerals received overall higher acceptance rates and, specifically, by the interaction pattern between numeral type and condition, according to which round numerals are accepted more in the within- and out-of-range conditions, but not in the false and matching conditions. These results show that when the divergence is small enough, addressees accept the use of a round, diverging numeral as reasonable. Thus, the idea underlying the preferred-approximation hypothesis, that addressees are aware of speakers' imprecise use of round numerals more than sharp ones, is supported by the findings.

In addition, the results from the pairwise comparisons of conditions show that participants were sensitive to the compatibility between the speaker's utterance and the factual information, with graded acceptability: Matching > Within range > Out of range > False. Generally, the greater the distance between the numeral in the utterance and the numeral in the state of affairs, the less participants accepted the utterance as a reasonable description of the state of affairs. That is, across addressees, divergence does not behave as a binary category, but rather as graded, with magnitude constituting the significant factor. Whether this is the case on the individual levels as well requires further examination.

Having tested the assumption that addressees recognize speakers' imprecise use of specifically round numerals, we now turn to testing the actual interpretation they derive for both types of numerals, in Experiments 2a, 2b, and 3.

3. Experiment 2a

Experiment 2a aims to tap into the speaker's intended meaning of numerals, focusing on precision. Participants were presented with sentences uttered by speakers which contained either round or sharp numerals. These sentences were presented under one of three conditions: modified by *exactly*, modified by *about*, or unmodified (the bare condition). Participants were then asked to accept or reject a statement explicating the speaker's sentence. This statement introduced either the same numeral as the sentence (the matching condition) or ± 3 range around that numeral (see **Figure 3** for a screen shot of an experimental trial).

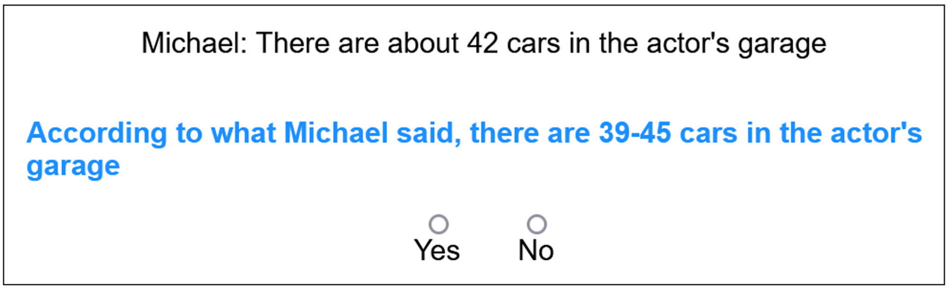

Figure 3: A screenshot of an experimental item in the sharp numeral condition in Experiment 2a.

3.1 Methods

3.1.1 Participants

240 native speakers of American English participated in Experiment 2a. Four participants were excluded, because they gave wrong answers to at least two filler items that had clear correct responses. The final number of participants was, thus, 236 (113 female, mean age = 31.3, *SD* = 5.52).

3.1.2 Materials

The materials consisted of stimulus sentences and target statements. The experiment followed a 3 × 2 × 2 design, with numeral type (round/sharp) manipulated between subjects, and both modifier (*exactly*/*about*/bare) and target type (number/range) manipulated within subjects. Each participant was exposed to one of 24 experimental lists, each containing 6 experimental items and 12 fillers. The order of items was randomized per participant.

We used the same base sentences and numerals as in Experiment 1, excluding 90. Thus, in the round numeral condition, the numerals were 20, 30, 40, 60, 70, and 80. In the sharp numeral condition, the same numerals were used plus 2 (e.g., 22, 32, etc.).

The numerals in the utterances appeared in three conditions: two control conditions (modified by *exactly* or *about*) and one critical condition (bare numeral, *e.g., There are 30 marbles…*).

Target statements appeared in one of two forms. In the number condition, the target statement included the same sentence and numeral as in the stimulus. In the range condition, the numeral was replaced with a ± 3 range around it (e.g., *27–33* for a stimulus sentence with *30*).

Each list included one version of every modifier × target type combination. Conditions were counterbalanced across lists.

The target sentences were prefaced with “According to what [SPEAKER] said.” We chose this phrasing because it allows participants to opt for what they take to be the speaker’s intended message, rather than for a verbatim repetition (such as “according to [SPEAKER]’s words”, or “[SPEAKER] said that…”).

Table 5 presents a single item in the round numeral condition across all six within-subject conditions (modifier × target type). All experimental items were rotated across these six conditions using a Latin Square design, so that each participant saw each combination only once, and each item appeared in all conditions across the full set of lists.

Table 5: A single stimulus item (*There are 30 marbles in the red bowl*) across all six within-subject conditions in the round numeral condition.

Modifier	Target Type	Stimulus Sentence	Target Sentence
Bare	Number	John: There are 30 marbles in the red bowl	According to what John said, there are 30 marbles in the red bowl
Bare	Range	John: There are 30 marbles in the red bowl	According to what John said, there are 27–33 marbles in the red bowl
Exactly	Number	John: There are exactly 30 marbles in the red bowl	According to what John said, there are 30 marbles in the red bowl
Exactly	Range	John: There are exactly 30 marbles in the red bowl	According to what John said, there are 27–33 marbles in the red bowl
About	Number	John: There are about 30 marbles in the red bowl	According to what John said, there are 30 marbles in the red bowl
About	Range	John: There are about 30 marbles in the red bowl	According to what John said, there are 27–33 marbles in the red bowl

12 additional filler items were included per numeral type: 3 with a numeral modified by *about*, 3 with numerals modified by *exactly*, and the rest with unmodified numerals. For round numerals, the filler items included multiples of 10 and 5 (e.g., *55*, *95*). In half of the fillers, the target sentence included a numeral; in the other half, it included a range. When the target included a numeral, that numeral was either the same, lower, or higher than the numeral in the stimulus. In some cases of range paraphrases, the ranges contained the numeral in the stimulus

(e.g., *63–67* for *about 65*), while in others, they did not (e.g., *83–87* for *70*). None of the fillers included the same range size as the critical items. This design ensured a roughly equal number of 'yes' and 'no' responses across the experiment, reducing response bias and encouraging genuine interpretive judgments, while still keeping numeral type as a between-subjects condition.

3.1.3 Procedure

During the experiment, participants read the target sentences and then accepted or rejected a statement about what happened, according to what the speaker said.

Before starting the experiment, participants were given two training trials to familiarize themselves with the task and to ensure they understood what the ranges stood for. In the first trial, the stimulus sentence contained a numeral modified by *exactly* (*exactly 97*), with a range in the target statement (*95–99*). If the participant rejected the statement, they received feedback that they were correct, along with the following explanation: "A range of numbers indicates approximation, namely, that the specific value is not pinpointed by the speaker. But [SPEAKER] did assert that there was a specific quantity of [ITEMS]". If they accepted the target statement, they were informed that they were incorrect, and the same explanation was presented to them.

In the second trial, the stimulus sentence contained a numeral modified by *about* (*about 90*), with a range in the target statement (*89–91*). If the participant rejected the target statement, they were told they were wrong, with the explanation: "A range of numbers indicates approximation, namely, that the specific value is not pinpointed by the speaker. Indeed, [SPEAKER] did not assert a specific quantity of [ITEMS]." If they accepted the target statement, they were told they were correct, and given the same explanation.

No example with a bare numeral was introduced during the training trials.

3.1.4 Predictions

The competing accounts about the interpretation of round numerals make distinct predictions regarding how participants should respond to different combinations of numeral type, modifier, and target type.

According to the preferred-approximation hypothesis, round numerals are more likely to be interpreted as approximate. Therefore, bare round numerals should lead to relatively high acceptance rates of range paraphrases. This does not mean that they should be exactly like those of numerals modified by *about*, as *about* is explicitly approximative, while for the round numerals, approximation results from an extra inferential step. Bare sharp numerals, by contrast, should pattern like those modified by *exactly*, leading to high acceptance of matching number paraphrases and low acceptance of ranges (again, the modification by an explicit *exactly* may trigger some difference). This view predicts a three-way interaction, where round numerals in the bare condition should receive higher acceptance rates when given range paraphrases than sharp ones.

According to the precise-by-default hypothesis**,** numerals are interpreted precisely by default. Therefore, bare numerals should pattern with numerals modified by *exactly* across both round and sharp numerals: both types of numerals should lead to high acceptance of number paraphrases, and low acceptance of range paraphrases. It does not mean that the acceptance rates would be identical, because *exactly* explicitly blocks imprecise interpretations, while for the bare numerals, precision is, according to the hypothesis, the default reading, but it can be overruled. This predicts no difference between round and sharp numerals in the bare condition.

Finally, we raised the possibility that there is no interpretive bias, or else that speakers vary in their interpretive strategies. Therefore, bare round numerals will elicit an inconsistent pattern, with intermediate or participant-specific variation. Round and sharp numerals might show similar mean acceptance rates across target types, but with greater variability. This predicts no consistent interaction effects.

These predictions are summarized in **Table 6**.

Table 6: Predicted acceptance of paraphrases in the *bare* condition.

Hypothesis	**Bare Round + Number**	**Bare Round + Range**	**Bare Sharp + Number**	**Bare Sharp + Range**
Preferred-approximation	NA	High	High	Low
Precise-by-default	High	Low	High	Low
Mixed strategy	Medium	Medium	Medium	Medium

3.1.5 Analyses

Data analysis followed the same Bayesian framework as in Experiment 1. We fit a Bayesian mixed-effects logistic regression predicting the binary response (yes/no) from numeral type (round/sharp), modifier (bare/about/exactly), target type (number/range), and their interactions. Factors were sum-coded.

Since maximal random-effects structures, including interaction slopes, led to convergence issues, we simplified them. We first removed the correlation parameters from the random-effects (Barr et al., 2013; Bates et al., 2015), but this did not improve convergence. We then simplified the random-effects structure by removing interaction slopes, while retaining random intercepts and slopes for all main effects. This approach ensures that we preserve the key sources of participant- and item-level variability, while minimizing overfitting and improving model stability (Matuschek et al., 2017).

We used weakly informative regularizing priors (Student-t(3, 0, 2.5) for fixed effects and variance components, and LKJ(2) for correlation matrices of fixed effects). Four MCMC chains were run for 8,000 iterations each (2,000 warmup), with adapt_delta = 0.99 and max_treedepth = 15 to ensure stable sampling. Regression coefficients are reported as posterior β estimates (medians) on the log-odds scale, with 95% credible intervals.

Planned pairwise comparisons were conducted using the emmeans package, and all contrasts are reported as posterior medians with 95% credible intervals.

3.2 Results

Table 7 presents the acceptance rates in the different conditions across both numeral types. Acceptance rates for *about N* with a range in the target, and bare N and *exactly N* with a number in the target, are at ceiling (recall that in the number condition, the same number appears in the stimulus and the target). When the stimulus numeral is either modified by *exactly* or bare and the target is a range, the acceptance rates plummet to around 10%. *About N* stimuli with a number target are accepted in about 30% of the trials.

Table 7: Results of Experiment 2a by modifier and target.

Modifier	Target	Acceptance rate	95% CIs
About	Number	.31	.25–.37
About	Range	.86	.81–.90
Bare	Number	.99	.97–1.00
Bare	Range	.10	.06–.13
Exactly	Number	.99	.97–1.00
Exactly	Range	.06	.03–.10

Next, we examine whether the response pattern differs between round and sharp numerals. The results by numeral type are presented in **Figures 4** and **5**. For both numeral types, numerals in the bare condition pattern with the *exactly N* condition, with high acceptance rates for number targets and extremely low acceptance rates in the range condition. For both numeral types, in the *about N* condition, responses are acceptable in high proportions for range targets (although less so for sharp numerals) and in medium proportions for number targets.

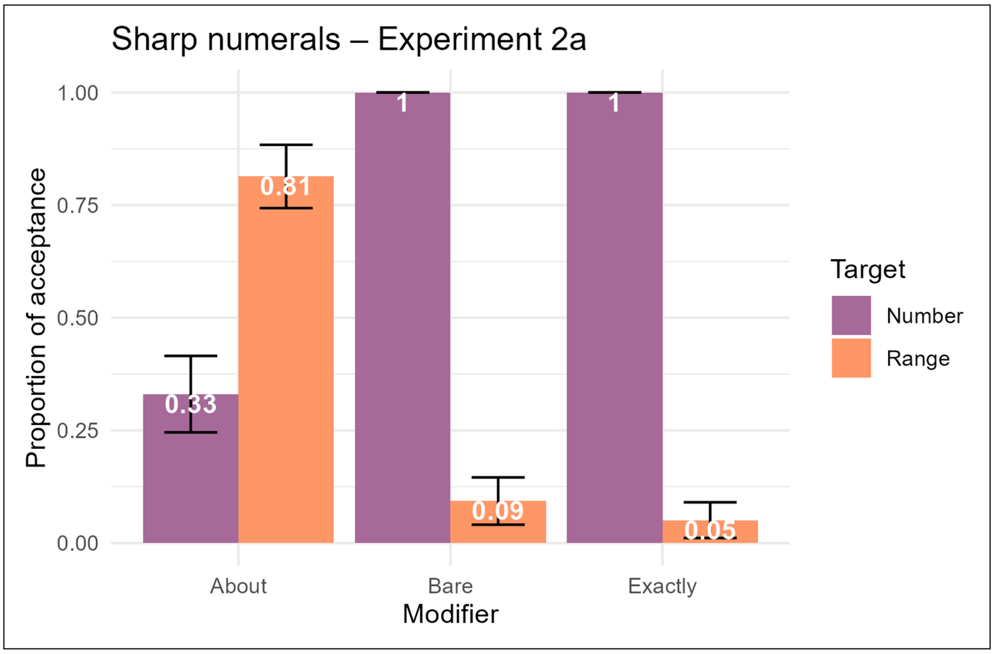

Figure 4: Results of Experiment 2a – sharp numerals. Error bars represent 95% CI.

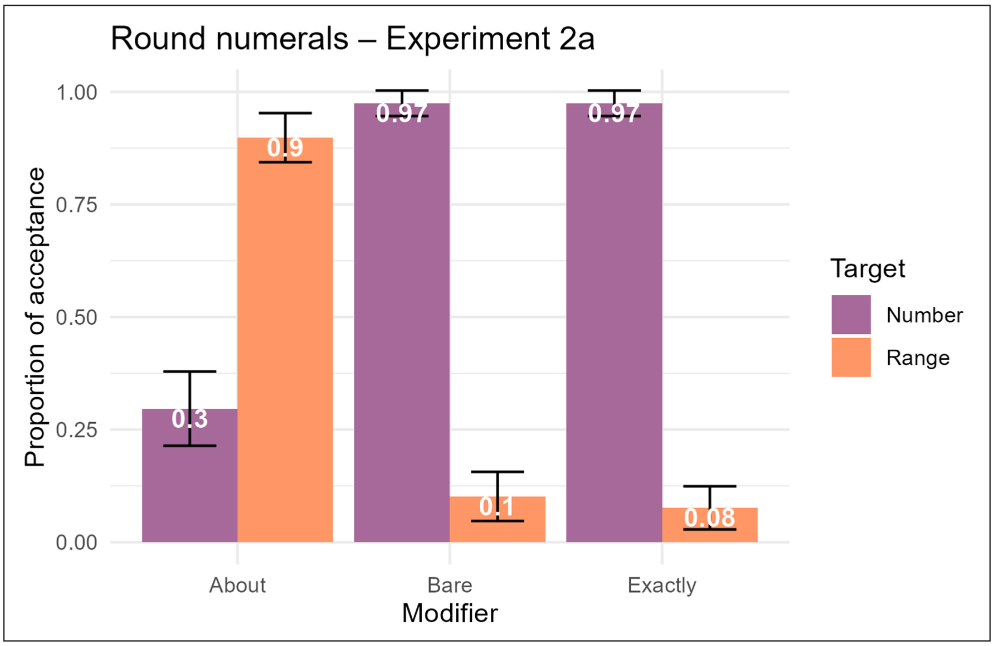

Figure 5: Results of Experiment 2a – round numerals. Error bars represent 95% CI.

The model's coefficients are presented in **Table 8**, and the model's predictions are visualized in **Figure 6**.

Table 8: Bayesian mixed-effects logistic model, formula: response ∼ type * modifier * target + (1 + type + modifier + target | participant) + (1 + type + modifier + target | item). Factors are sum-coded. Therefore, the intercept reflects the grand mean.

Predictor	Estimate	Std. Error	95% CI	$\hat{R}$
Intercept (Grand Mean)	2.96	1.06	[1.31, 5.39]	1.00
Type: Round	–0.99	0.69	[–2.48, 0.23]	1.00
Modifier: About	–1.61	0.83	[–3.46, –0.20]	1.00
Modifier: Bare	0.47	0.80	[–1.06, 2.12]	1.00
Target: Number	5.28	1.51	[3.02, 8.84]	1.00
Round × About	1.52	0.70	[0.37, 3.07]	1.00
Round × Bare	–0.49	0.69	[–1.94, 0.85]	1.00
Round × Number	–1.39	0.62	[–2.80, –0.36]	1.00
About × Number	–8.31	2.27	[–13.75, –5.15]	1.00
Bare × Number	2.90	1.04	[1.19, 5.35]	1.00
Round × About × Number	0.73	0.59	[–0.31, 2.03]	1.00
Round × Bare × Number	0.12	0.68	[–1.28, 1.46]	1.00

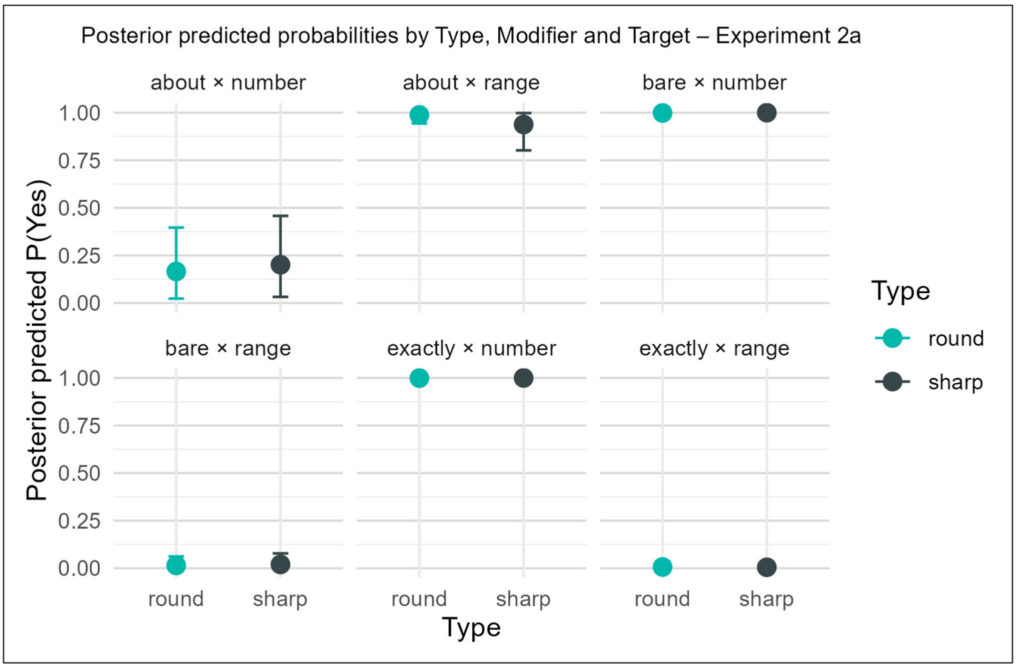

Figure 6: Posterior predicted probabilities of acceptance in Experiment 2a for round and sharp numerals, separated by Modifier and Target. Points represent posterior medians; error bars represent 95% HPD credible intervals.

The model showed a credible main effect of modifier, driven by lower acceptance in the *about* condition (β = –1.61, 95% CI [–3.46, –0.20]). There was also a credible main effect of target type, with number targets yielding higher acceptance than range targets (β = 5.28, 95% CI [3.02, 8.84]). In contrast, the main effect of numeral type was not credible (β = –0.99, 95% CI [–2.48, 0.23]).

However, because both modifier and target appeared in credible two-way interactions between modifier and target and between numeral type and target, the main effects are not interpreted in isolation. The three-way interaction among type, modifier, and target showed no credible evidence (β = 0.73, 95% CI [–0.31, 2.03]; β = 0.12, 95% CI [–1.28, 1.46]).

To understand the interaction between modifier and target, we conducted pairwise comparisons. For targets containing numbers, the acceptance rates were higher in the bare and *exactly* conditions than in the *about* condition. For targets containing ranges, the results were reversed, with trials in the *about* condition leading to higher acceptance rates than in both the *exactly* and bare conditions. Full pairwise comparisons appear in **Table 9**.

Table 9: Pairwise comparisons of modifier by target type.

Target Condition	Contrast	Estimate	95% CI
Number	About – Bare	–12.67	[–21.07, –6.99]
	About – Exactly	–15.81	[–26.49, –7.53]
	Bare – Exactly	–2.96	[–9.66, 3.06]
Range	About – Bare	8.55	[5.30, 14.37]
	About – Exactly	10.26	[5.90, 17.82]
	Bare – Exactly	1.64	[–0.85, 5.02]

As for the target type and numeral type interaction, for number targets, acceptance was slightly higher in the sharp condition than in the round condition. For range targets, the opposite pattern was observed: round numerals were accepted more than sharp ones. However, the difference was only credible for number targets. Full pairwise comparisons are presented in **Table 10**.

Table 10: Pairwise comparisons of numeral type by target type.

Target Condition	Contrast	Estimate	95% CI
Number	Round – Sharp	–4.53	[–9.59, –0.44]
Range	Round – Sharp	0.72	[–1.4, 3.01]

3.3 Discussion

Overall, we see that bare numerals, both round and sharp, lead to high acceptance rates of number paraphrases (97%–100% acceptance) and rejection of range paraphrases (9%–10% acceptance). This pattern is similar to that of *exactly* modified numerals. These results are not consistent with the predictions of the preferred-approximation theory, which expects round numerals to favor range interpretations. They are also inconsistent with the weaker version of the hypothesis, according to which both range and number paraphrases should be moderately accepted. Instead, they are more consistent with the prediction of the precise-by-default proposal, which predicts high acceptance of number paraphrases and high rejection of range paraphrases. Importantly, the same pattern is observed for both numeral types (although there is a small difference in the number condition — see below). We note that participants did not have to choose one type of interpretation over the other, and, in principle, could confirm both number and range paraphrases. Despite that, they completely rejected the range paraphrases for both bare numeral types and, crucially, for round numerals.

At the same time, we also observed a trend in which number targets were slightly more accepted for sharp numerals, and range targets for round numerals. Although the actual differences between round and sharp numerals are small – around 3% across modifiers (see **Figures 4–5**), this pattern may suggest that round numerals are slightly more compatible with imprecise interpretations than sharp ones. The trend in the range condition, though not credible, may suggest that sharp numerals resist coercion into imprecise interpretations. These possibilities warrant further testing. Importantly, greater compatibility does not entail that round numerals are, by default, interpreted imprecisely, as the behavioral pattern here shows that, in this task, in both numeral types, participants rejected the range interpretations. Therefore, this finding is consistent with the precise-by-default hypothesis, as it generates predictions about interpretation rather than compatibility.

Now, there are two concerns about the results of Experiment 2a. The first, and more substantial, one is that the ranges in the paraphrases might not correspond to participants' actual interpretations of round numerals. This is because the ranges were defined quite strictly, and participants may have inferred a different range, either narrower or wider, around the numeral. Moreover, approximate interpretations do not require a specific range to be derived. It is possible that these interpretations come with an unspecified, fuzzy meaning. In both cases, participants may have rejected the range paraphrases not because they interpreted the numerals precisely, but because the paraphrase did not match their inferred approximation.

A second concern raised by a reviewer is that the training phase may have encouraged participants to reject range paraphrases for round numerals, as they may have understood from the training that such paraphrases are only acceptable when approximation is explicitly marked. To address both concerns, we devised Experiment 2b, which was identical to Experiment 2a, except for two changes. First, instead of a specific range, the range paraphrases were replaced with *close to N*, allowing for an approximate interpretation without specifying numeric boundaries. Second, the training phase was removed entirely.

4. Experiment 2b

4.1 Method

4.1.1 Participants

240 native speakers of American English participated in Experiment 2a. Six participants were excluded, because they gave wrong answers to at least two filler items that had clear correct responses. The final number of participants was, thus, 234 (117 female, mean age = 43.14, *SD* = 13.94).

4.1.2 Materials

Same as Experiment 2a, only this time, range paraphrases were changed such that instead of the range, the paraphrase included *close to N*.[8] **Figure 7** presents an example of an experimental item.

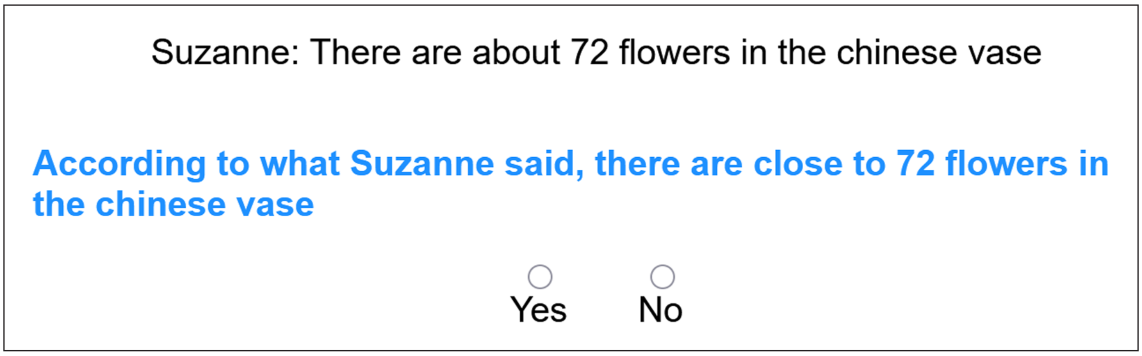

Figure 7: A screenshot of an experimental item in the sharp numeral condition in Experiment 2b.

[8] One of our reviewers notes that *close to* may trigger an inference that the speaker does not believe the exact value holds. Thus, the rejection of *close to* paraphrases in the *exactly* condition may reflect this inference, rather than the speaker's uncertainty about the precise value. While this possibility cannot be entirely ruled out, it does not affect the interpretation of the bare condition, on which the main theoretical predictions bear. Moreover, the reviewer's (reasonable) objection is to a speaker uttering *close to N*, but *close to* is only an attempt to describe the interpretation. It is, therefore, not judged as an actual utterance that must be cooperative.

4.1.3 Procedure

Same as Experiment 2a, only without the training phase.

4.1.4 Predictions

Same as Experiment 2a.

4.1.5 Analyses

Same as Experiment 2a.

4.2 Results

Table 11 presents the acceptance rates in the different conditions across both numeral types for Experiments 2a and 2b, for comparability. Acceptance rates for *about N* with *close to* in the target, and *bare N* and *exactly N* with a number in the target, remain at ceiling. In the rest of the trials, acceptance rates in Experiment 2b are higher. This is especially evident in the bare + range/*close to* conditions, which rose from 10% to 37%, and in the *exactly* + range/*close to* conditions, which increased from 6% to 21%. Conversely, participants were much more accepting of number paraphrases for numerals modified by *about*, as evidenced by the fact that acceptance in the *about* + number condition was also higher in Experiment 2b (68% vs. 31%).

Table 11: Results of Experiments 2a and 2b by modifier and target.

Modifier	Target	Acceptance (Exp 2b)	95% CIs (Exp 2b)	Acceptance (Exp 2a)
About	Number	.68	.62–.74	.31
About	Range/close to	.87	.83–.91	.86
Bare	Number	.99	.97–1	.99
Bare	Range/close to	.37	.3–.43	.10
Exactly	Number	.99	.98–1	.99
Exactly	Range/close to	.21	.16–.26	.06

The results by numeral type are presented in **Figures 8** and **9**. Here too, for both numeral types, numerals in the bare condition pattern with the *exactly N* condition, with high acceptance rates for number targets, only now the acceptance rates in the range condition are moderately low. For both numeral types, in the *about N* condition, responses are acceptable in high proportions for both target types.

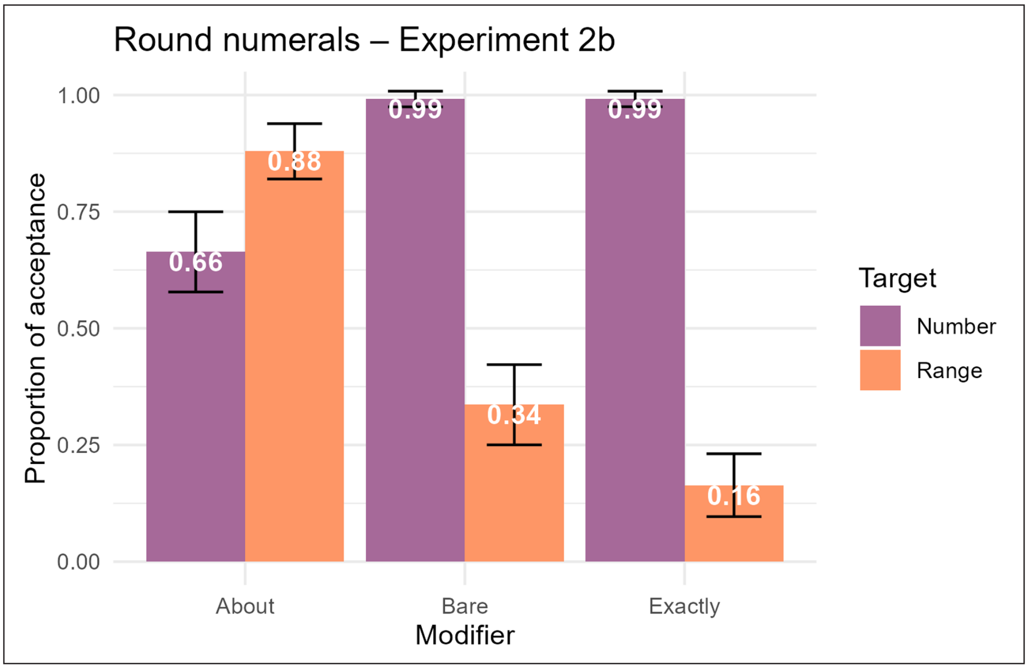

Figure 8: Results of Experiment 2b – sharp numerals. Error bars represent 95% CI.

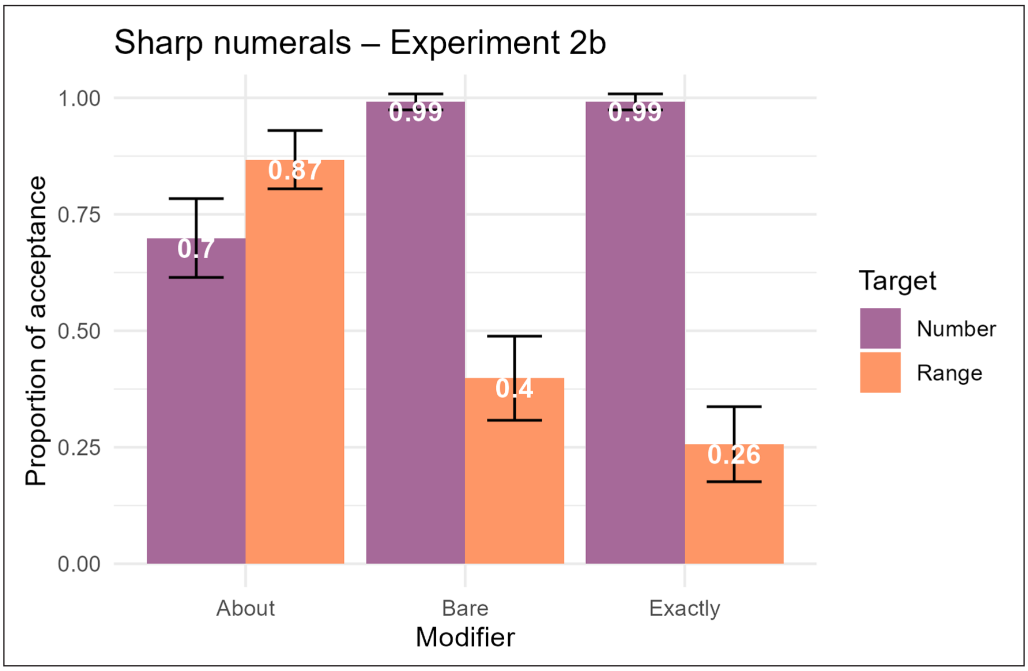

Figure 9: Results of Experiment 2b – round numerals. Error bars represent 95% CI.

The model's coefficients are presented in **Table 12**, and the model's predictions are visualized in **Figure 10**.

Table 12: Bayesian mixed-effects logistic model, formula: response ~ type * modifier * target + (1 + type + modifier + target | participant) + (1 + type + modifier + target | item). Factors are sum-coded. Therefore, the intercept reflects the grand mean.

Predictor	Estimate	Std. Error	95% CI	$\hat{R}$
Intercept (Grand Mean)	5.69	1.94	[2.83, 10.32]	1.00
Type: Round	–0.23	0.62	[–1.50, 0.96]	1.00
Modifier: About	–2.15	1.04	[–4.56, –0.53]	1.00
Modifier: Bare	0.86	0.74	[–0.40, 2.54]	1.00
Target: Number	6.54	2.23	[3.26, 11.88]	1.00
Round × About	0.29	0.55	[–0.75, 1.46]	1.00
Round × Bare	–0.24	0.56	[–1.42, 0.83]	1.00
Round × Number	0.16	0.53	[–0.89, 1.24]	1.00
About × Number	–8.16	2.69	[–14.57, –4.37]	1.00
Bare × Number	1.74	1.00	[0.16, 4.08]	1.00
Round × About × Number	–0.44	0.52	[–1.54, 0.51]	1.00
Round × Bare × Number	–0.17	0.55	[–1.33, 0.89]	1.00

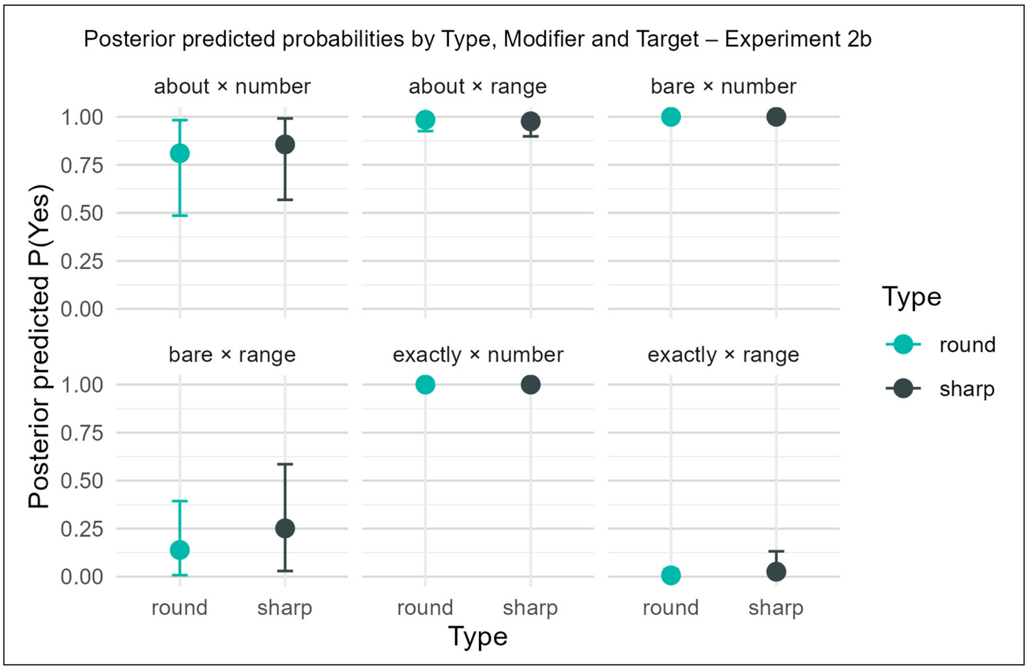

Figure 10: Posterior predicted probabilities of acceptance in Experiment 2b for round and sharp numerals, separated by Modifier and Target. Points represent posterior medians; error bars represent 95% HPD credible intervals.

As in Experiment 2a, the model showed a credible main effect of modifier, driven by lower acceptance in the *about* condition relative to the grand mean (β = –2.15, 95% CI [–4.56, –0.53]). There was also a credible main effect of target type, with number targets leading to higher acceptance than *close to* targets (β = 6.54, 95% CI [3.26, 11.88]). The main effect of numeral type (β = –0.23, 95% CI [–1.50, 0.96]) and the three-way interaction among type, modifier, and target were not credible (β = –0.44, 95% CI [–1.54, 0.51]; = –0.17, 95% CI [–1.33, 0.89]).

As for the modifier × target interaction, pairwise comparisons show that sentences containing *about* led to high acceptance of both number and *close to* paraphrases. In contrast, for bare numerals and *exactly*, participants strongly accepted number paraphrases, but rejected *close to* paraphrases. While the difference between bare numerals and *exactly* numerals is not credible in the number condition, *exactly* led to lower acceptance rates than bare in the *close to* condition. The full pairwise comparisons are presented in **Table 13**.

Table 13: Pairwise comparisons of modifier by target in Experiment 2b.

Target	Contrast	Estimate	95% CI
Number	**About – Bare**	–12.17	[–21.88, –5.74]
	About – Exactly	–16.84	[–31.87, –7.52]
	Bare – Exactly	–4.62	[–13.39, 1.42]
Close to	**About – Bare**	6.36	[3.34, 11.83]
	About – Exactly	10.31	[5.26, 18.96]
	Bare – Exactly	3.87	[1.31, 8.41]

Finally, although the type x target interaction term did not reach significance, we conducted a pairwise comparison to see if we could detect a trend similar to the one found in Experiment 2a. In both conditions, round numerals elicited lower acceptance rates (unlike Experiment 2a, where round numerals were accepted more in the range condition), but the contrasts were not credible. The full pairwise comparisons are presented in **Table 14**.

Table 14: Pairwise comparisons of type by target in Experiment 2b.

Target Condition	Contrast	Estimate	95% CI
Number	Round – Sharp	–0.14	[–4.33, 3.79]
Close to	Round – Sharp	–0.72	[–2.94, 1.40]

4.3 Discussion

The results found no credible interaction of numeral type with any other factor. Both round and sharp numerals followed the same patterns as before, showing high acceptance for number paraphrases and much lower acceptance for *close to* paraphrases, except in the *about* condition. Additionally, the weak trend between numeral type and target type observed in Experiment 2a was not replicated here, reinforcing the conclusion that, in this task, there was no evidence for roundness licensing an approximate interpretation over sharpness.

Bare numerals of both types aligned more closely with the behavior of *exactly*-modified numerals than with *about*: In the *exactly* condition, acceptance of number paraphrases is at ceiling, which is also the case for bare numerals. Furthermore, both types show a dispreference for *close to* paraphrases, although bare numerals show higher acceptance rates of those paraphrases, as is evident from the pairwise comparisons of modifier by target. This contrasts sharply with the *about* condition, which shows a qualitatively different pattern: higher acceptance for *close to* paraphrases than for number paraphrases.

We note that the changes made from Experiment 2a resulted in Experiment 2b showing generally higher acceptance rates for approximate paraphrases compared to Experiment 2a, and this difference is mostly driven by higher acceptance of *close to* paraphrases in the bare and *exactly* conditions. It's difficult to assess whether this increase has to do with *close to* replacing the ranges, or with the removal of the practice phase in Experiment 2b, which, in Experiment 2a, had instructed participants to reject range paraphrases for *exactly*-modified sentences.

5. Experiment 3

While Experiments 2a and 2b probed the speaker-intended *interpretation* of numerals, Experiment 3 tested the *compatibility* of the different types of numerals with imprecise interpretations in a truth judgment task. Even though Experiments 2a and 2b suggested that round numerals were interpreted as precise, a weaker version of the preferred-approximation hypothesis was not ruled out. It's possible that round numerals can be taken as more compatible with approximation than sharp ones, even if their default interpretation is precise.

To test this, we used a TVJ task in which participants were presented with a sentence containing a round or a sharp numeral, followed by a fact about the actual numerosity. Participants were asked to judge whether the speaker's utterance was right. Traditionally, this task has been used in semantics and pragmatics studies to tap into the semantic meaning of utterances, because meaning was equated with truth conditions, and the latter were supposed to provide the (sole) basis for the truth judgment. This assumption has been challenged, however, by recent studies, according to which truth judgment is a more "liberal" task, allowing for some discrepancy between the speaker-intended message and the actual state of affairs, due to the application of truth-compatible inferences (Ariel, 2004) or pragmatic tolerance (Katsos & Bishop, 2011). Indeed, Shetreet et al. (in prep.) argue that TVJ tasks trigger different results from what they call participant-control tasks, as in Experiments 2a and 2b: Specifically, they show that TVJ manifests a higher tolerance for lower-bounded only interpretations for Hebrew 'some' expressions, as well as ad hoc scalars (see also Fishman et al., 2023, for *or*). Thus, we use this task to test whether participants will, indeed, mobilize one of the above principles, which will then allow them to judge a number range as compatible with round numerals more than with sharp ones.

If round numerals are more easily seen as compatible with different values (pointing to a possible approximate interpretation) than sharp ones are, we should see an increase in the acceptance rates of utterances with round numerals where the numeral in the fact diverges slightly from the numeral in the utterance, as compared with utterances containing sharp numerals.

5.1 Method

5.1.1 Participants

420 native speakers of American English (209 females, mean age = 42, SD = 12.9) participated in Experiment 3.

5.1.2 Materials

The experiment followed the same 2 × 4 design as Experiment 1, with numeral type (round/sharp) manipulated between subjects and condition (matching/within-range/out-of-range/ false) manipulated within subjects. The materials were the same as in Experiment 1.

5.1.3 Procedure

Same as Experiment 1, but this time participants were asked to make a binary judgment about whether the speaker was right, given the fact. **Figure 11** presents an example of a trial.

5.1.4 Analyses

The analysis followed the same Bayesian framework as Experiment 1. The number of iterations was increased to 6000 (2000 warmup), after preliminary models with fewer iterations showed signs of convergence issues (low effective sample sizes for some parameters). As in Experiment 1, we modeled response as a binary variable (yes/no) as a function of numeral type (round/sharp), condition (matching/within range/out of range/false), and their interaction. The model included random slopes and intercepts for participants and items. Categorical predictors were sum-coded.

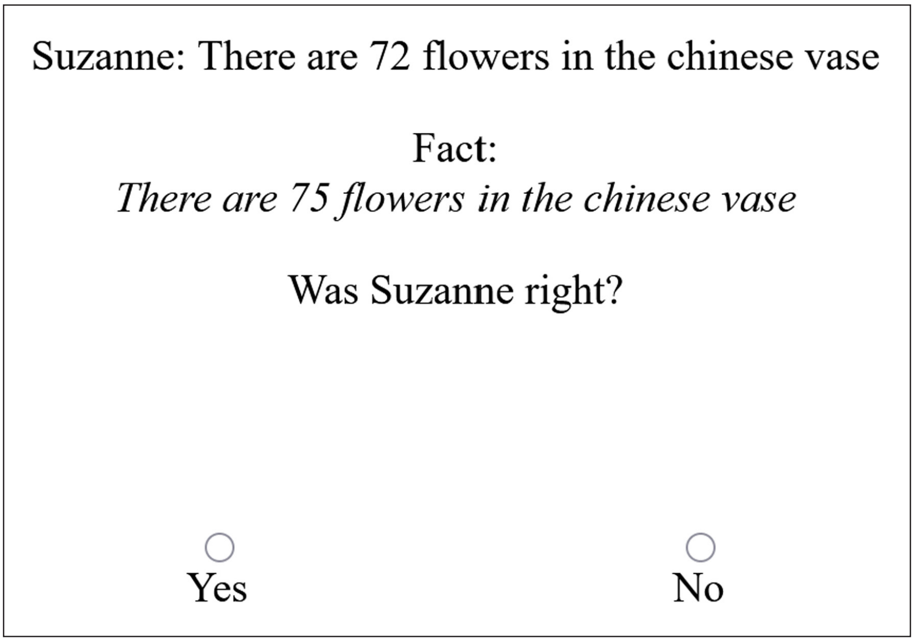

Figure 11: A screenshot of a trial from Experiment 3.

5.2 Results

The results of Experiment 3 are presented in **Figure 12**.

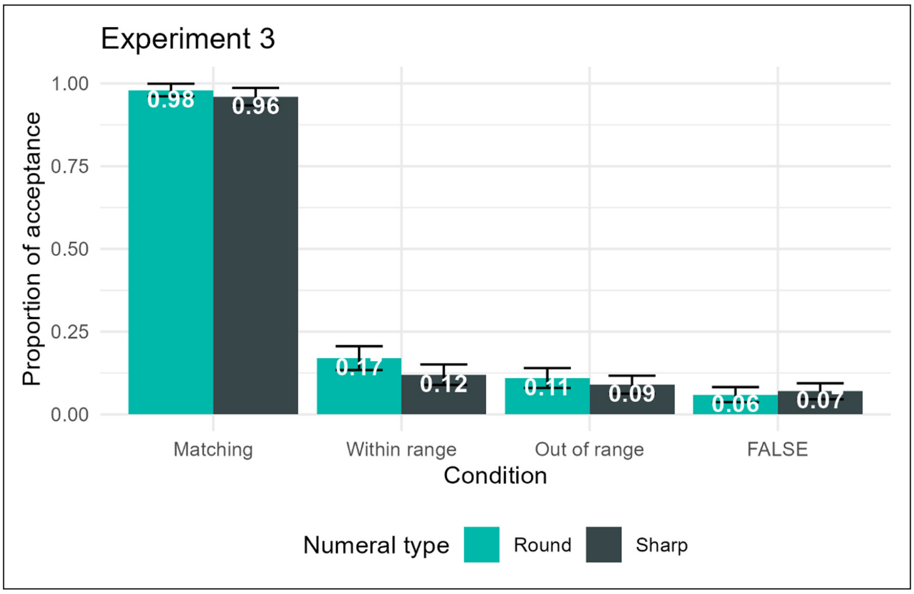

Figure 12: Proportion of acceptance of the speaker's utterance by numeral type and condition in Experiment 3. Error bars stand for 95% CIs.

The model coefficients are presented in **Table 15**, and the model's predictions are visualized in **Figure 13**.

Table 15: Bayesian mixed-effects logistic model, formula: response ∼ type * condition + (1 + type * condition | participant) + (1 + type * condition | item). Factors are sum-coded. Therefore, the intercept reflects the grand mean.

Predictor	Estimate	Std. Error	95% CI	$\hat{R}$
Intercept (Grand Mean)	–1.74	0.28	[–2.29, –1.21]	1.00
Type: Round	0.25	0.23	[–0.22, 0.72]	1.00
Condition: False	–2.71	0.36	[–3.46, –2.06]	1.00
Condition: Matching	7.35	0.70	[6.17, 8.93]	1.00
Condition: Out of Range	–2.13	0.33	[–2.82, –1.54]	1.00
Round × False	–0.29	0.34	[–0.95, 0.41]	1.00
Round × Matching	0.57	0.53	[–0.36, 1.72]	1.01
Round × Out of Range	–0.20	0.26	[–0.75, 0.28]	1.00

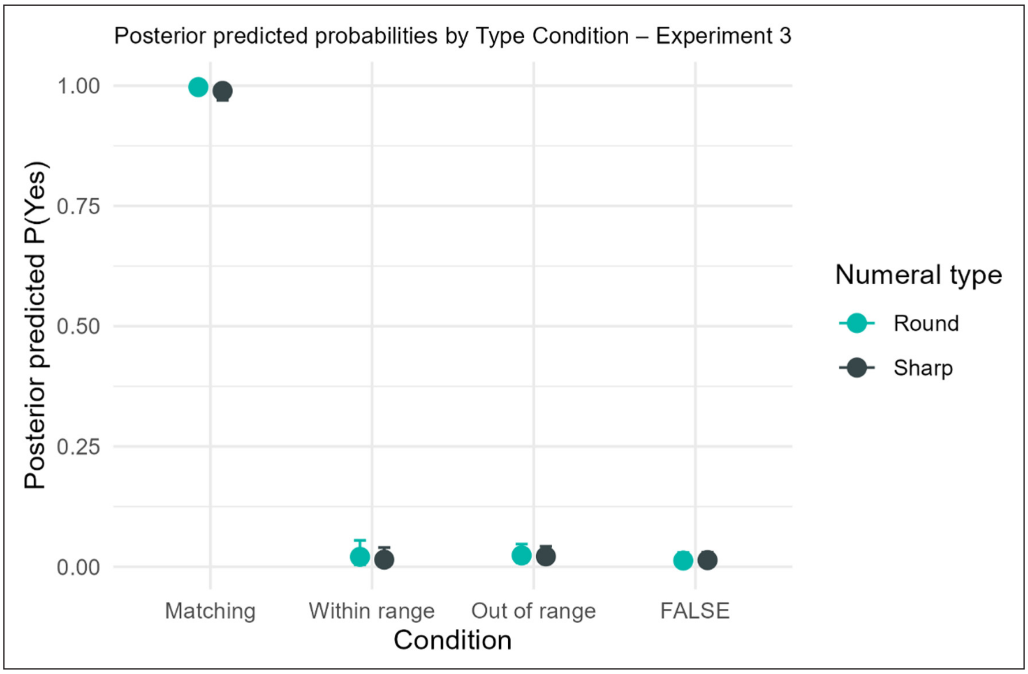

Figure 13: Posterior predicted probabilities of a 'Yes' response by numeral type and condition. Points represent the posterior mean estimates from the Bayesian mixed-effects model: response ∼ type * condition + (1 + type * condition | participant) + (1 + type * condition | item). Error bars represent 95% Highest Posterior Density (HPD) credible intervals.

The model found a credible effect of condition, with matching receiving higher acceptance rates relative to the grand mean (β = 7.35, 95% CI [6.17, 8.93]). The main effect of numeral type was not credible (β = 0.25, 95% CI [–0.22, 0.72]). The interaction between type and condition showed no credible evidence (β = –0.29, 95% CI [–0.95, 0.41], β = 0.57, 95% CI [–0.36, 1.72], β = –0.20, 95% CI [–0.75, 0.28]).

Table 16 presents the acceptance rates and CIs for each condition. Acceptance was highest in the Matching condition, followed by Within range, with Out of range and False conditions showing much lower acceptance. These results replicate the graded acceptability pattern found in Experiment 1, despite the different task. However, not all contrasts were credible in a pairwise comparison (**Table 17**).

Table 16: Acceptance rates and 95% CIs of the four conditions in Experiment 3.

Condition	Acceptance rate	95% CIs
Matching	.97	.95–.99
Within range	.15	.12–.17
Out of range	.10	.08–.12
False	.07	.05–.09

Table 17: Pairwise comparisons of condition in Experiment 3.

Contrast	Estimate	95% CI
False – Matching	–9.98	[–11.96, –8.20]
False – Out of range	–0.58	[–1.28, 0.12]
False – Within range	–0.21	[–1.37, 0.98]
Matching – Out of range	9.39	[7.71, 11.40]
Matching – Within range	9.78	[7.99, 11.85]
Out of range – Within range	0.38	[–0.66, 1.46]

5.3 Discussion

Not surprisingly, participants overwhelmingly accepted the speaker's utterance as true in the matching condition. However, given the robust intuition in the literature about round numerals being naturally interpretable as approximate, one might expect participants to be more willing to accept as "right" the use of round numerals when the actual value is close to it, even if it slightly diverges from the numeral. However, the results did not support this expectation, suggesting that even minor discrepancies between the numeral and the actual value led to strong rejections. Our results found substantial rejections of non-matching values in all non-matching conditions. Even the within-range condition showed only a 15% acceptance rate.

We note that our findings differ from previous findings regarding the potential association of round numerals with approximate values. For example, as many as 35% (or even 45%) of Beltrama and Schwarz's (2022) participants confirmed as "Right" the use of a round numeral (e.g., 200) for a state of affairs of a sharp numeral (e.g., 207.06) (the different rates are a function involving the persona of the speaker). Several methodological and conceptual differences between the studies may contribute to this contrast. Most importantly, our experiments tested *counting numerals*, whereas Beltrama and Schwarz examined *measurement numerals*, which have been argued to license approximate construals more readily (Ariel & Levshina, 2024; Cummins, 2015). Their stimuli also included decimal values, which are quite marked and fall outside the basic number counting scale, potentially encouraging approximate interpretations. Taken together, these differences may help explain why our results show lower overall acceptance of approximate readings.

Now, we have so far assumed that the basis for judging true a stimulus numeral describing a state of affairs where an in-range number diverges from the stimulus number points to the acceptance of a compatibility between the actual state of affairs and the stimulus numeral. The expectation is that, if this is the case, we should perhaps see this more often with round numerals. However, our results found no evidence for that being the case.[9] However, three potential issues remain. The first concerns the asymmetry between trials in which the numeral in the fact is below the sentence numeral as opposed to above it. The second concerns the decision to include all conditions in the analysis. These issues are addressed in 5.4 and 5.5, respectively. The third concerns the epistemic state attributed by the participants to the speakers. This is addressed in the supplementary files available at the OSF site for this article.

5.4 Position of the target numeral with respect to the sentence numeral

Two theoretical accounts predict an asymmetry between higher-number and lower-number divergences, i.e., when *30* is said to describe a state of affairs of '33' versus '27'. Both suggest higher acceptance rates for higher than for lower values. The originally consensual account posits a lower-bound only ('at least') semantic meaning for numerals, with exact interpretations derived via a quantity implicature (Gazdar, 1979; Grice, 1975; Horn, 1972). According to this

[9] We note that this finding does not only argue against preferred-approximation for round numerals. It also argues against an underspecified semantic meaning for numerals (Carston, 1998; Geurts, 2006). An underspecified meaning should be able to accommodate in-range divergences, and possibly more so for round numerals.

analysis, since higher values are part of the numeral's meaning, using a numeral when the actual number is higher should be more acceptable than when the actual number is lower.

Others have argued that the meaning (or the "what-is-said" interpretation) of numerals is circumbounded, i.e., lower and upper-bounded, and is, therefore, exact (Ariel, 2004; Breheny, 2008; Horn, 1992; Koenig, 1991). Still, such an analysis is not incompatible with higher confirmations for higher-number divergences. Ariel (2004) predicts that a truth-compatible inference can bridge the gap, between e.g., '30' and a state of affairs of '33', because if there are, e.g., 33 chairs, then there are 30 chairs. Higher values do not contradict the speaker's commitment to '30' (the speaker might have been under-informative), but lower values (e.g., '27') do, making it more difficult to infer compatibility.

We should, then, check this hypothesis. Note that the prediction here applies equally to round and sharp numerals. If, however, we find higher confirmations of cases in which the numeral in the fact is below the uttered numeral for round numerals than for sharp numerals, this could be an argument in favor of approximation, because neither account can explain such a difference. 5.4.1 reports our post hoc analysis, which investigates whether rounding is influenced by whether the number in the fact was below or above the numeral in the sentence.

5.4.1 Analysis

We performed a similar Bayesian analysis on the data of Experiment 3, but excluded the matching condition (note that while the matching condition is present in the figures, it was excluded from the analysis). We used weakly informative priors: Normal(0, 2.5) priors for fixed effects, a wider Normal(0, 5) prior for the intercept, Student-t(3, 0, 2.5) priors for random-effects standard deviations, and an LKJ(2) prior for correlation matrices.

The model's fixed effects included numeral type, condition (within range, out of range, and false), and the additional predictors of the position of the number in the fact (above or below the utterance) and the interaction term between all three predictors. For the random-effects structure, we followed the same procedure as in Experiments 2a and 2b. The model included an additive random-effects structure for participants and items. (following the same procedure as in Experiments 2a and 2b). Factors were sum-coded.

5.4.2 Results

Figures 14 and **15** present the results of Experiment 3 again, this time with acceptance rates for non-matching conditions separated into fact numerals that are higher than the utterance numeral and fact numerals that are lower than the utterance numeral.

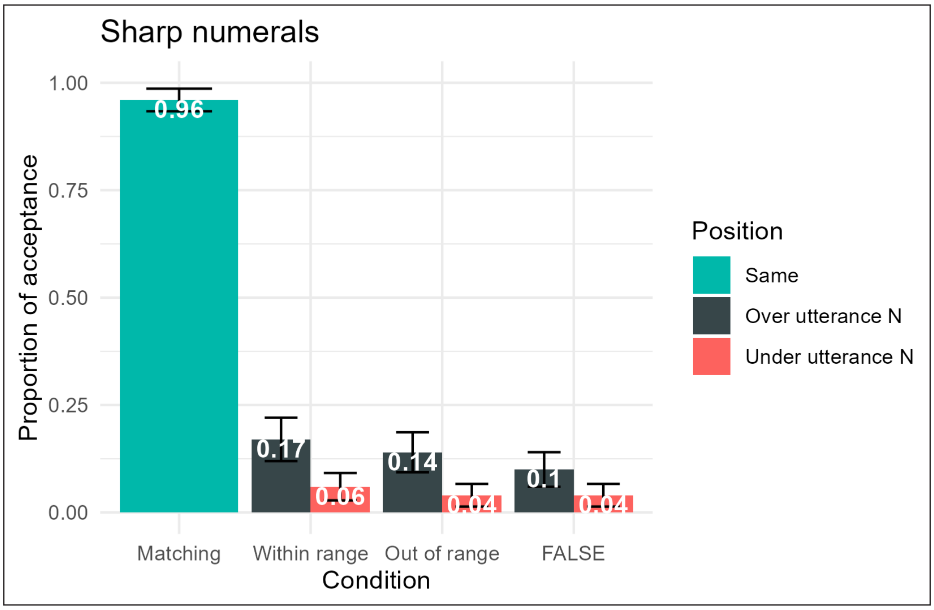

Figure 14: Results of Experiment 3, including acceptance rates by the type of numeral in the fact, for sharp numerals.

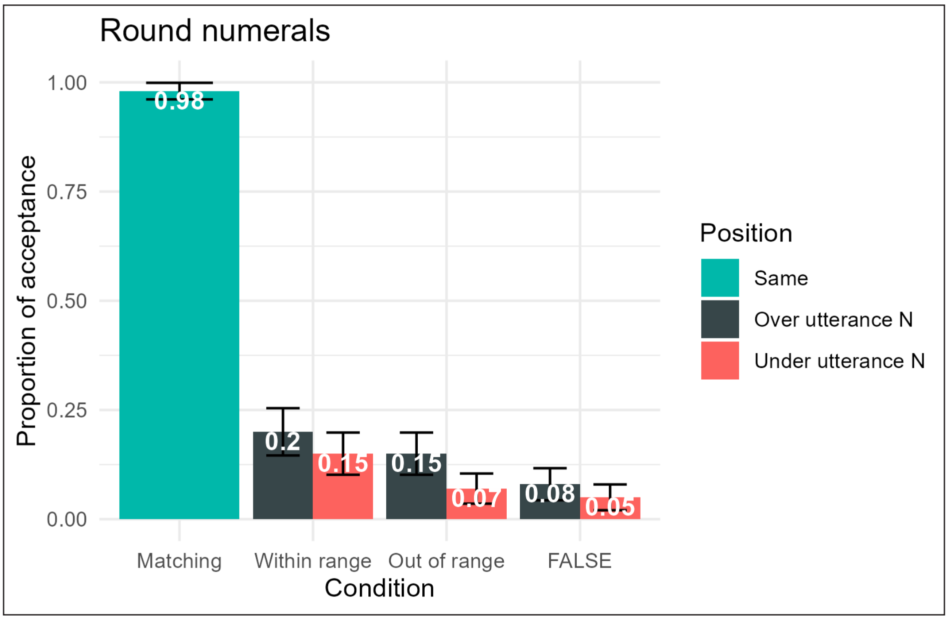

Figure 15: Results of Experiment 3, including acceptance rates by the type of numeral in the fact, for round numerals.

The model's coefficients are presented in **Table 18**, and the model's predictions are presented in **Figure 16**.

Table 18: Bayesian mixed-effects logistic model, formula: response ~ type * condition * position + (1 + type + condition + position | participant) + (1 + type + condition + position | item). Factors are sum-coded. Therefore, the intercept reflects the grand mean.

Predictor	Estimate	Std. Error	95% CI	$\hat{R}$
Intercept (Grand Mean)	–11.14	2.15	[–16.00, –7.68]	1.00
Type: Round	0.68	0.61	[–0.48, 1.96]	1.00
Condition: False	0.10	0.76	[–1.42, 1.62]	1.00
Condition: Out of Range	0.19	0.61	[–1.04, 1.42]	1.00
Position: Over	1.45	0.75	[0.09, 3.01]	1.00
Round × False	–0.47	0.52	[–1.45, 0.62]	1.00
Round × Out of Range	–0.17	0.41	[–1.05, 0.58]	1.00
Round × Over	–0.81	0.49	[–1.91, 0.02]	1.00
False × Over	–0.80	0.44	[–1.74, –0.02]	1.00
Out of Range × Over	0.33	0.34	[–0.31, 1.04]	1.00
Round × False × Over	–0.01	0.31	[–0.64, 0.58]	1.00
Round × Out of Range × Over	0.24	0.27	[–0.27, 0.79]	1.00

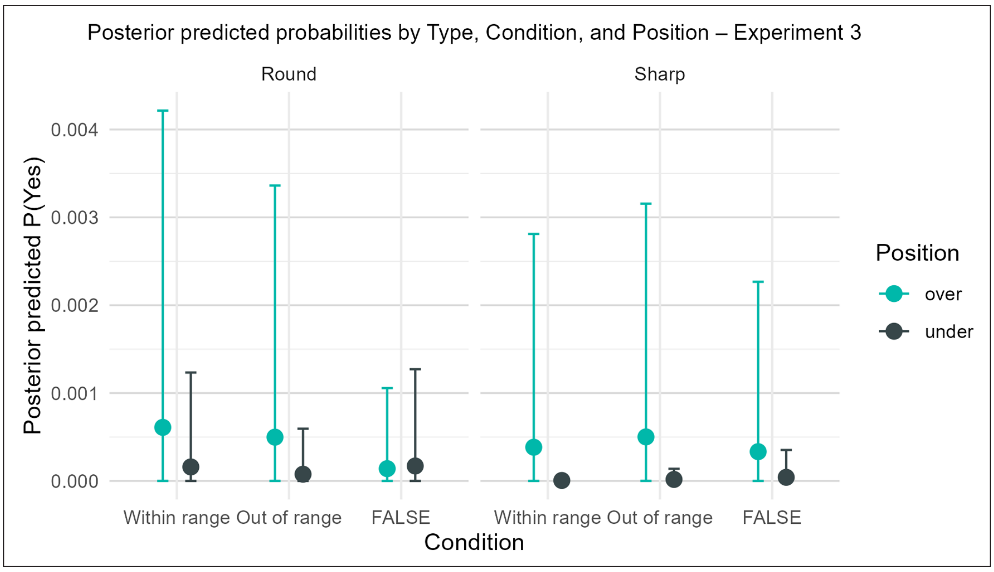

Figure 16: Posterior predicted probabilities of a 'Yes' response by numeral type and condition. Points represent the posterior mean estimates from the Bayesian mixed-effects model: response ~ type * condition * position + (1 + type + condition + position | participant) + (1 + type + condition + position | item). Error bars represent 95% Highest Posterior Density (HPD) credible intervals.

The analysis revealed a credible main effect of position (β = 1.45, 95% CI [0.09, 3.01]), with higher acceptance rates when the fact N was larger than the utterance N. There was no credible main effect of numeral type (β = 0.68, 95% CI [–0.48, 1.96]), nor evidence of a credible interaction between type and position or condition, or for the triple interaction.

The model also found a credible interaction between numeral condition and position (β = –0.80, 95% CI [–1.74, –0.02]). However, pairwise comparisons found no credible differences (see **Table 19**).

Table 19: Pairwise comparisons of condition by position.

Position	Contrast	Estimate	95% CI
Above	False – Out of range	–1.16	[–3.59, 1.05]
	False – Within range	–0.86	[–3.87, 2.00]
	Out of range – Within range	0.29	[–1.98, 2.83]
Below	False – Out of range	0.98	[–1.66, 4.05]
	False – Within range	1.56	[–1.81, 5.25]
	Out of range – Within range	0.58	[–2.37, 3.70]

5.4.3 Discussion

As predicted by the two accounts of the meaning of numerals mentioned above, factual values above the utterance N were more acceptable than those below it (7% vs. 14%). Note, however, that overall, even the higher diverging values received quite low acceptance rates.

Importantly, the interaction between type and position was not found to be credible. Round numerals show a higher acceptance rate of values below the round numeral than sharp ones do (15% acceptance rate for round numerals, 6% for sharp ones, more than twice as much). This could have been interpreted as an effect of rounding: while sharp numerals are overwhelmingly incompatible with values below them, round numerals may (albeit rather infrequently) license a compatibility inference when the fact numeral is lower than the numeral in the speaker's message. In our data, however, this tendency is marginal. Non-matching values are still rejected at very high rates for both numeral types, and the overall pattern offers little support for the idea that round numerals are more compatible with an approximate interpretation than a precise interpretation, or that there is no preference. Since the study was not designed to examine values below the reference numeral, further research would be needed to determine how robust this effect truly is.

5.5 Focusing on within-range trials

A possible concern raised by a reviewer is that the predicted effect of numeral type on acceptance of approximate trials should be assessed specifically within the *within-range* condition, rather than across all conditions, as this is where differences in interpretation are most likely to surface. Additionally, acceptability might vary with numerical magnitude: because the range of acceptable values was fixed across all numerals, the proportion of divergence differed. For example, a 3-unit difference from 20 represents about 15%, whereas the same difference from 60 is only 5%. If participants are sensitive to proportional divergence, we might expect lower acceptance for numerals with larger relative deviations.

We therefore conducted a focused analysis including only within-range trials. We fit a Bayesian logistic regression model using brms's package's default weakly informative priors: Normal(0, 1) priors for fixed effects, a Student-t(3, 0, 2.5) prior for the intercept, and a Student-t(3, 0, 2.5) prior for the random intercept standard deviation. We modeled response as a function of numeral type (round/sharp), position (above/below), scaled numeral magnitude (z-scored), and their interactions as fixed effects.

A model with a maximal random-effects structure failed to converge. We therefore applied the same stepwise simplification procedure as in the previous analysis. Because an additive random-effects structure did not resolve the convergence problems, we progressively removed random-effects terms with the smallest estimated variances and refit the model at each step. This procedure continued until convergence was achieved, resulting in a model that included only random intercepts for participants. More complex random-effects structures consistently exhibited convergence problems, manifested in inflated intercept estimates (greater than 20). The model's coefficients are presented in **Table 20**.

Table 20: Bayesian mixed-effects logistic model, formula: Response ~ type*position + (1|participant). Factors are sum-coded. Therefore, the intercept reflects the grand mean.

Predictor	**Estimate**	**Std. Error**	**95% CI**	$\hat{R}$
Intercept (Grand Mean)	–3.76	0.47	[–4.78, –2.95]	1.00
Type: Round	0.47	0.23	[0.03, 0.93]	1.00
Magnitude	0.24	0.17	[–0.08, 0.57]	1.00
Position: Over	0.68	0.16	[0.38, 1.01]	1.00
Round × Magnitude	0.13	0.17	[–0.20, 0.46]	1.00
Round × Over	–0.32	0.15	[–0.63, –0.03]	1.00
Magnitude × Over	–0.09	0.18	[–0.45, 0.27]	1.00
Round × Magnitude × Over	–0.12	0.18	[–0.49, 0.24]	1.00

The analysis revealed credible main effects of type ($\beta = 0.47$, 95% CI [0.03, 0.93]), with higher acceptance rates for round numerals, and position ($\beta = 0.68$, 95% CI [0.38, 1.01]), with higher acceptance rates when the fact N was larger than the utterance N. No credible main effect of numerical magnitude was found ($\beta = 0.24$, 95% CI [–0.08, 0.57]).

The interaction between type and position was credible (β = –0.32, 95% CI [–0.63, –0.03]), with follow-up contrasts showing that round numerals were more acceptable than sharp ones only in the *under* condition (estimate = 1.58, 95% CI [0.41, 2.78]), with no difference in the *above* condition (estimate = 0.29, 95% CI [–0.66, 1.26]).

The interactions of Magnitude × Type, Magnitude × Position, and the three-way interaction were not credible.

Unlike the analysis in 5.4.2, this analysis found credible evidence that round numerals are slightly more compatible with approximation than sharp numerals, specifically when the fact numeral was smaller than the sentence numeral ('below' condition). However, despite this relative difference, acceptance rates for round numerals remained low overall, a finding which does not indicate that they are inherently compatible with approximation in this task, and that they are only less incompatible than sharp numerals. Finally, the magnitude of the numerals had no measurable effect on the acceptability of approximate interpretations. But in view of the magnitude effect on preference for round numerals reported in Woodin et al. (in prep.), it may very well be that the magnitude difference between our small and large numbers is just not large enough.

6. General discussion and concluding remarks

The two main findings of this study are that: (a) addressees are aware of the speakers' license to *use* round numerals imprecisely (Experiment 1); however, (b) we found no evidence for a preference for approximate *interpretations* of round numerals. Sharp and round numerals alike consistently showed patterns that match a precise interpretation (Experiments 2a, 2b), even in the more "tolerant" task (Experiment 3).

Considering the competing theoretical analyses of round numerals, our results fail to find strong support for the preferred-approximation hypothesis. Recall that the preferred-approximation proposal claimed that round numerals should favor approximate over precise interpretations. The evidence we found for Krifka's (2002, 2007) proposal is the difference between round and sharp numerals in the within-range condition in Experiment 2a (about 3% higher acceptance for round numerals), and the increased acceptance of round numerals when a state of affairs contained a higher numeral (15% versus 5%, which was only found credible in the analysis in 5.5). Experiments 2b and 3 found no credible interactions between numeral type and any other predictor. These results are more in line with the predictions of the precise-by-default hypothesis. Additionally, in experiments 2a and 2b, which test interpretation, we found that bare round numerals led to acceptance of number paraphrases and rejection of range paraphrases, a pattern similar to that of numerals modified by *exactly*.

Still, the intuition, widely shared by linguists and non-linguists alike, that round numerals are more easily associated with imprecise values is partially supported by Experiment 1. The picture that emerges is that of a mismatch between speakers and addressees. Speakers may choose to use round numerals imprecisely (see also Woodin et al., in prep.), and addressees, moreover, recognize this (Experiment 1), but, nonetheless, we found little evidence that they actually interpret round numerals as approximate (Experiments 2a, 2b, and 3).

The findings of Experiments 2a, 2b, and 3 were quite surprising, since we found only very weak evidence for approximate responses to round numerals. These findings are surprising for two reasons. First, they contradict the robust intuitions about rounding, as well as claims based on experiments in the literature. Even in the precise-by-default hypothesis, there is no claim that round numerals should not be compatible with an approximate interpretation, even though Experiment 3 found no such compatibility. The second reason is that Experiment 1 found that addressees find it more reasonable to use round numerals than sharp ones for imprecise descriptions. Pragmatic theory generally posits that interlocutors mentalize each other's intentions and knowledge states. We would, then, expect that if addressees recognize that the speaker may be rounding, they would at least be more charitable in a TVJ task. However, this is not the case.

We briefly discuss two potential explanations for this pattern. The first is relevance-theoretic (Sperber & Wilson, 1986), according to which, while speakers may be motivated to prefer round over sharp numerals for the various reasons introduced in 1.1, this speaker preference may not dictate an addressee's engagement with these motivations when decoding the speaker's message. Addressees may, instead, settle for a default interpretation that adequately serves the conversation's goals, perhaps in order to keep their own processing costs to a minimum. Thus, they fail to consider the speaker's preferences and motivations. This entails a speaker/addressee asymmetry.

A second possible explanation is that an imprecise description of the facts on the speaker's part does not automatically entail a *communicative intention* to actually convey imprecision or approximation. It is possible that the addressee does not derive an approximate interpretation for round numerals, because approximation is not, in fact, part of the speaker's *intended message*. Instead of assuming that addressees ignore the speaker's perspective, we may focus on the mismatch between the state of affairs (as perceived by the speaker, say, '37' or '43') and her choice of *message* (and, hence, her utterance, say, *40*). Even when the speaker knows the precise number, she may, nonetheless, prefer a message that is easier for her to produce and for the addressee to process (a round numeral). Once the speaker has chosen a certain message, the addressee is bound by it. The addressee doesn't overthink, they needn't engage in reverse engineering. Unless there are specific contextual (or linguistic) indicators to the contrary, the addressee assumes that the speaker intended them to interpret the numeral according to its default (precise) meaning. In other words, addressees do not ignore the speaker's intentions and preferences, but rather assume that these preferences are for the precise interpretation. On this account, there is no mismatch between speakers and their addressees, which is in line with standard assumptions about cooperative communication.

The experiments presented here do not allow us to decide between the two explanations, and we leave it to future research to determine which is a better explanation for the gap between the results of Experiment 1, on the one hand, and Experiments 2a, 2b, and 3, on the other hand. But whichever explanation holds, the gap between our Experiment 1 results (speakers are licensed to round off numbers) and experiments 2a, 2b, and 3 (addressees may not actually interpret round numerals as imprecise) should caution linguists to at least not automatically assume that a speaker's general preference for round numbers, such as the one found in Mühlenbernd and Solt (2022), Van Der Henst et al. (2002), and Woodin et al. (in prep.), even if recognized by the addressee, translates into an actual imprecise interpretation by the addressee.

Data accessibility statement

The data and scripts used for the analysis in this study are available at: https://osf.io/3qrjt/overview?view_only=75896122fc06479e97703dc51d0bcbb5.

Supplementary files, which include Appendices A and B, are available in the same repository.

Ethics and consent

This study was approved by the Ethics Committee at Tel Aviv University. All participants provided informed consent prior to participation.

Acknowledgments

We thank all those who have engaged with earlier versions of this work. We are also grateful to our anonymous reviewers and the handling editor, Florian Schwarz, for their valuable feedback.

Funding information

This research was supported by the Israel Science Foundation (grant no. 1398/20 to Mira Ariel).

Competing interests

The authors have no competing interests to declare.

Author contributions

Nicole Katzir: Conceptualization, Formal analysis, Methodology, Resources (stimuli generation), Data curation, Writing – original draft, Writing – review & editing

Mira Ariel: Conceptualization, Methodology, Resources (stimuli generation), Writing – review & editing, Supervision

ORCiD IDs

Nicole Katzir: https://orcid.org/0000-0001-7024-800X

Mira Ariel: https://orcid.org/0000-0003-2018-6917

References

Anscombre, J.-C., & Ducrot, O. (1989). Argumentativity and informativity. In M. Meyer (Ed.), *From metaphysics to rhetoric* (pp. 71–87). Kluwer. https://doi.org/10.1007/978-94-009-2593-9_6

Aparicio, H. (2017). *Processing context-sensitive expressions: The case of gradable adjectives and numerals* [Doctoral dissertation]. University of Chicago.

Ariel, M. (2004). Most. *Language*, *80*(4), 658–706. https://doi.org/10.1353/lan.2004.0162

Ariel, M. (2006). A 'just that' lexical meaning for *Most*. In K. von Heusinger & K. Turner (Eds.), *Where semantics meets pragmatics* (pp. 49–91). Elsevier. https://doi.org/10.1163/9780080462608_007

Ariel, M. (2017). Ad hoc-ization of numeral categories. *Paper presented at Building categories in interaction: Multidisciplinary approaches to categorization*. University of Bologna.

Ariel, M. (2019). Numerous puzzles. Paper presented at *ZAS workshop on The meaning of numerals: Cognitive, experimental, and semantic perspectives*. ZAS, Berlin.

Ariel, M. (2021). Why it's hard to construct ad hoc number concepts. In C. Mauri, I. Fiorentini, & E. Goria (Eds.), *Building categories in interaction: Linguistic resources at work* (pp. 439–462). John Benjamins. https://doi.org/10.1075/slcs.220.16ari

Ariel, M., & Levshina, N. (2024). The counting principle makes number words unique. *Corpus Linguistics and Linguistic Theory*, *21*(1), 173–199. https://doi.org/10.1515/cllt-2023-0105

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357. https://doi.org/10.1037/0096-1523.10.3.340

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bastiaanse, H. 2011. The Rationality of Round Interpretation. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), Vagueness in Communication (Lecture Notes in Computer Science), 37–50. Berlin, Heidelberg: Springer. DOI: https://doi.org/10.1007/978-3-642-18446-8_3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beltrama, A., & Schwarz, F. (2022). Imprecision, personae, and pragmatic reasoning. *Semantics and Linguistic Theory*, *31*, 122. https://doi.org/10.3765/salt.v31i0.5107

Breheny, R. (2008). A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics*, *25*(2), 93–139. https://doi.org/10.1093/jos/ffm016

Bürkner, P.-C. (2021). Bayesian item response modeling in *R* with brms and Stan. *Journal of Statistical Software*, *100*(5). https://doi.org/10.18637/jss.v100.i05

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 179–236). Benjamins. https://doi.org/10.1075/pbns.37.11car

Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Blackwell. https://doi.org/10.1002/9780470754603

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenonenon. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics. An international handbook of natural language meaning* (pp. 2297–2332). De Gruyter. https://doi.org/10.1515/9783110253382.2297

Coupland, N. (2011). How frequent are numbers? *Language & Communication*, *31*(1), 27–37. https://doi.org/10.1016/j.langcom.2010.09.001

Cummins, C. (2015). *Constraints on numerical expressions*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199687909.001.0001

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29. https://doi.org/10.1016/0010-0277(92)90030-L

Dorogovtsev, S. N., Mendes, J. F. F., & Oliveira, J. G. (2006). Frequency of occurrence of numbers in the World Wide Web. *Physica A: Statistical Mechanics and Its Applications*, *360*(2), 548–556. https://doi.org/10.1016/j.physa.2005.06.064

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., Englebretson, R., & Martey, N. (2000). *Santa Barbara corpus of spoken American English, Parts 1–4*. Linguistic Data Consortium.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Ferson, S., O'Rawe, J., Antonenko, A., Siegrist, J., Mickley, J., Luhmann, C. C., Sentz, K., & Finkel, A. M. (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

Fishman, A., Asherov, D., Katzir, N., & Ariel, M. (2023). Discourse context cannot make 'or' inclusive (only experimental task can). *Journal of Pragmatics*, *217*, 7–16. https://doi.org/10.1016/j.pragma.2023.08.017

Gazdar, G. (1979). *Pragmatics, implicature, presupposition and logical form*. Academic Press.

Gega, P., Liu, M., & Bechberger, L. (2021). Numerical concepts in context. In L. Bechberger, K.-U. Kühnberger, & M. Liu (Eds.), *Concepts in action, language, cognition, and mind* (pp. 93–119). Springer. https://doi.org/10.1007/978-3-030-69823-2_5

Geurts, B. (2006). Take 'five': The meaning and use of a number word. In S. Vogeleer & L. Tasmowski (Eds.), *Indefiniteness and plurality* (pp. 311–329). Benjamins. https://doi.org/10.1075/la.95.16geu

Grice, P. H. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics. Vol. 3: Speech Acts* (pp. 41–58). Academic Press. https://doi.org/10.1163/9789004368811_003

Horn, L. (1972). *On the semantic properties of logical operators in English* [doctoral dissertation]. UCLA.

Horn, L. (1989). *A natural history of negation*. University of Chicago Press.

Horn, L. (1992). The said and the unsaid. *Ohio State University Working Papers in Linguistics (SALT II Proceedings)*, *40*, 163–192. https://doi.org/10.3765/salt.v2i0.3039

Jansen, C. J. M., & Pollmann, M. M. W. (2001). On round numbers: Pragmatic aspects of numerical expressions. *Journal of Quantitative Linguistics*, *8*(3), 187–201. https://doi.org/10.1076/jqul.8.3.187.4095

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), 12002–12007. https://doi.org/10.1073/pnas.1407479111

Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81. https://doi.org/10.1016/j.cognition.2011.02.015

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*(1), 1–45. https://doi.org/10.1007/s10988-006-9008-0

Koenig, J.-P. (1991). Scalar predicates and negation: Punctual semantics and interval interpretations. *Chicago Linguistic Society*, *27*, 140–155.

Krifka, M. (2002). Be brief and vague! And how Bidirectional Optimality Theory allows for verbosity and precision. In D. Restle & D. Zaefferer (Eds.), *Sounds and systems. Studies in structure and change. A festschrift for Theo Vennemann* (pp. 439–458). Mouton de Gruyter. https://doi.org/10.1515/9783110894653.439

Krifka, M. (2007). Approximate interpretations of number words: A case for strategic communication. In G. Bouma, I. Kramer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 111–126). Royal Netherlands Academy of Arts and Sciences.

Lasersohn, P. (1991). Pragmatic halos. *Language*, *75*(3), 522–551. https://doi.org/10.2307/417059

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means*. https://CRAN.R-project.org/package=emmeans

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press. https://doi.org/10.7551/mitpress/5526.001.0001

Lotz, J. (1955). On language and culture. *International Journal of American Linguistics*, *21*(2), 187–189. https://doi.org/10.1086/464329

Mason, J. D., Healy, A. F., & Marmie, W. R. (1996). The effects of rounding on memory for numbers in addition problems. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *50*(3), 320–323. https://doi.org/10.1037/1196-1961.50.3.320

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mühlenbernd, R., & Solt, S. (2022). Modeling (im)precision in context. *Linguistics Vanguard*, *8*(1), 113–127. https://doi.org/10.1515/lingvan-2022-0035

Nuerk, H., Iversen, W., & Willmes, K. (2004). Notational modulation of the SNARC and the MARC (Linguistic Markedness of Response Codes) effect. *The Quarterly Journal of Experimental Psychology Section A*, *57*(5), 835–863. https://doi.org/10.1080/02724980343000512

Odic, D., & Starr, A. (2018). An introduction to the approximate number system. *Child Development Perspectives*, *12*(4), 223–229. https://doi.org/10.1111/cdep.12288

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281. https://doi.org/10.1080/17470216508416445

Paradis, C. (2001). Adjectives and boundedness. *Cognitive Linguistics*, *12*(1). https://doi.org/10.1515/cogl.12.1.47

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Ronderos, C. R., Noveck, I., & Lossius Falkum, I. (2024). Straight enough: Deriving imprecise interpretations of maximum standard absolute adjectives. *Glossa Psycholinguistics*, *3*(1). https://doi.org/10.5070/G60111411

Shetreet, E., Tehan, T., Shabanov, Y., & Ariel, M. (in preparation). *Taking control over weak scalar expressions*. Tel Aviv University.

Sigurd, B. (1988). Round numbers. *Language in Society*, *17*(2), 243–252. https://doi.org/10.1017/S0047404500012781

Solt, S., Cummins, C., & Palmović, M. (2017). The preference for approximation. *International Review of Pragmatics*, *9*(2), 248–268. https://doi.org/10.1163/18773109-00901010

Solt, S., & Stevens, J. (2018). Some three students: Towards a unified account of *some*. *Semantics and Linguistic Theory*, *28*, 345. https://doi.org/10.3765/salt.v28i0.4436

Sperber, D., & Wilson, D. (1986). *Relevance*. Blackwell.

Van Der Henst, J., Carles, L., & Sperber, D. (2002). Truthfulness and relevance in telling the time. *Mind & Language*, *17*(5), 457–466. https://doi.org/10.1111/1468-0017.00207

Woodin, G., Lorson, A., & Winter, B. (in preparation). *Approximate round number use depends on magnitude*. University College London, University of Groningen, University of Birmingham.

Woodin, G., Winter, B., Littlemore, J., Perlman, M., & Grieve, J. (2023). Large-scale patterns of number use in spoken and written English. *Corpus Linguistics and Linguistic Theory, 20*(1), 123–152. https://doi.org/10.1515/cllt-2022-0082

Wu, Y., & Aparicio, H. (2025). Disagreements do not automatically raise the standard of precision. In T. Knowlton, F. Schwarz, & A. Papafragou (Eds.), *ELM 3* (pp. 435–446). Linguistic Society of America. https://doi.org/10.3765/elm.3.5835

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832