Round Numbers Can Sharpen Cognition

Huy A. Nguyen
hn1@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Jake M. Hofman
jmh@microsoft.com
Microsoft Research
New York, New York, USA

Daniel G. Goldstein
dgg@microsoft.com
Microsoft Research
New York, New York, USA

ABSTRACT

Scientists and journalists strive to report numbers with high precision to keep readers well-informed. Our work investigates whether this practice can backfire due to the cognitive costs of processing multi-digit precise numbers. In a pre-registered randomized experiment, we presented readers with several news stories containing numbers in either precise or round versions. We then measured their ability to approximately recall these numbers and make estimates based on what they read. Our results revealed a counter-intuitive effect where reading round numbers helped people better approximate the precise values, while seeing precise numbers made them worse. We also conducted two surveys to elicit individual preferences for the ideal degree of rounding for numbers spanning seven orders of magnitude in various contexts. From the surveys, we found that people tended to prefer more precision when the rounding options contained only digits (e.g., "2,500,000") than when they contained modifier terms (e.g., "2.5 million"). We conclude with a discussion of how these findings can be leveraged to enhance numeracy in digital content consumption.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

number representation; round numbers; measurement; experimentation; cognitive effect; recall; estimation

ACM Reference Format:
Huy A. Nguyen, Jake M. Hofman, and Daniel G. Goldstein. 2022. Round Numbers Can Sharpen Cognition. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3491102.3501852

1 INTRODUCTION

The digital age has brought about a significant volume of information which people access every day. According to a recent study, between 2016 and 2018, Americans on average consumed nearly 8 hours of media per day from sources including television, social media apps, and web browsing [3]. An important skill in making sense of this information is numeracy, the ability to process basic numerical concepts [14]. Unfortunately, varying degrees of innumeracy pose difficulties to both the producers and consumers of numerical information [2, 11, 22, 38, 39]. Regarding this issue, we share the same perspective with [18] – that lack of numeracy is primarily due to the representation of the data, rather than the capability of the reader. As people turn increasingly towards *digital* information sources, the HCI community can play a role in ameliorating the difficulties associated with numerical cognition, by building intelligent digital interfaces that automatically make numerical information easier to represent and understand.

Recent HCI research on improving numerical and statistical comprehension has proposed a wide range of solutions, ranging from novel tools to large-scale empirical comparisons of information and visualization formats [10, 24, 28, 29, 41]. Prior work has utilized external artifacts, including new types of plots and personalized analogies, to help users better contextualize the numerical information at hand. However, less attention has been paid to an even more fundamental question: how do number representations themselves impact people's comprehension? In this work, we focus on two ways of representing numbers: the precise format (e.g., 3,792) and the round format (e.g, 4,000). Given that round numbers are easier to process [8, 48], but also inherently less precise [31], we are interested in the cognitive consequences, either positive or negative, of replacing precise numbers with rounded ones. Specifically, can decreasing the precision of information shown to people increase their accuracy on cognitive tasks, and if so, can we leverage these insights to build intelligent interfaces that automatically make digitized information easier to understand?

In this work, we conducted three studies to investigate the effects of displaying round numbers and people's preferences for seeing round numbers. First, in a large-scale pre-registered experiment, we tested whether showing people round numbers improves their performance in comprehension tasks. To do so, we randomized whether people saw precise or round versions of numbers in news stories, then measured their ability to recall those numbers and make estimates based on what they read. A key aspect of our approach is that *we always measured accuracy based on the precise values* (and not on their rounded versions), *regardless of the condition people were assigned to.* For example, if the precise version of an article contained the number 3,792 and the rounded version showed it as 4,000, then accuracy would be measured relative to 3,792 in *both* conditions. This put participants in the round condition at an apparent disadvantage because they were shown less precise information; however, we found that the benefits of seeing round numbers outweigh the costs: people who saw rounded versions of numbers were more than 20 percentage points more likely to recall numbers close to the precise values in the articles, compared to those who actually saw the precise values.

Having demonstrated the benefits of showing round numbers, we then conducted two surveys to understand people's preferences for the ideal degree of rounding they would like to see in various

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3501852

contexts. In both of these studies, we showed people example news quotes with precise numbers and asked them to evaluate the balance of precision and readability of all possible rounding options for those numbers. In one version of the survey, when showing numbers in the millions and billions we used written modifiers (e.g., "58.4 million"), whereas in the other version we used only digits without any modifier words (e.g., "58,400,000"). We found that, when shown modifiers, people generally preferred to see numbers expressed with two to three non-zero leading digits, regardless of the magnitude of the number. Interestingly, however, this preference changed when people were shown the digit-only representations, in which case they preferred to see more non-zero leading digits as the magnitude of the number increases.

In the remainder of the paper, we provide a summary of our literature review on related topics, followed by detailed descriptions, results and discussions of each experiment and survey.

2 RELATED WORK

Round numbers are conventionally defined in terms of their divisibility properties. Dehaene and Mehler [16] described round numbers as either multiples of 10 or integer divisors of 60, while Jansen and Pollmann [26] examined whether a number can be expressed as a single digit multiple of 10^n, 2×10^n or 5×10^n - the more of these properties a number has, the rounder it is. Alternatively, Krifka [31] suggested that roundness is dependent not only on a number's mathematical attributes, but also on the scale of granularity in consideration. For instance, when referring to time, 2:45 may be considered rounder than 2:40 because 15-minute units are more commonly used in time measurement than 5-minute units. In the scope of this work, we will focus on numerical values that are frequently encountered in daily news and readings (e.g., stock price, population count, inflation rate), whose roundness property is more aligned with the power-of-ten interpretation by [26].

To summarize existing work on round numbers, we start by describing theories and empirical findings around the effects they are known to induce. Then we outline conventions around rounding numbers in different contexts. Finally, we review prior tools that assist with number comprehension to better understand the design and usability aspects that may emerge.

2.1 Effects of round numbers

Studies in communication have shown that, across several languages and contexts, round numbers are used much more frequently than non-round numbers of similar magnitudes [16, 23, 26]. A common explanation for this phenomenon involves the Relevance Theory [46], which posits that human cognition tends to be geared towards maximum relevance. Here relevance is determined in terms of cognitive benefit and processing effort; highly relevant information should provide positive cognitive effects while requiring low processing cost. In this view, round numbers have the potential to be relevant, given that they can convey roughly the same information as their precise counterparts while being easier to process [49].

Several studies have investigated Relevance Theory and its implications on round numbers, in the context of time information exchange [19, 45, 49]. An example finding from such studies is that, when asked about time, people tend to provide rounded answers to the nearest 5-minute or 15-minute unit, even when they have a digital watch that displays the exact time, in order to produce the more relevant utterance and reduce the inquirer's processing effort [49]. These results are extended by Solt and colleagues [45], who demonstrated that the round answer time formats are indeed more beneficial to the information receiver. In their study, participants who worked with round clock times demonstrated better and faster recall and manipulation of these time values than those who worked with the precise times.

A future work direction which Solt et al. proposed [45], and which we address in our study, is whether this finding generalizes to numerical information beyond clock times. To the best of our knowledge, there has been limited research in this area. The most related study was conducted by [33], in which participants were given addition problems of the form "current assets + noncurrent assets = total assets" and were told they would be tested on their memory of the total assets. The total assets were five-digit numbers that were displayed in either round format (with two digits followed by three zeros) or non-round format (with two digits followed by three non-zero digits). Participants completed a 90-second distraction task and were then given 90 seconds to write down all the sums they could recall. In addition, they were asked to identify, among a list of given numbers, which number was also a sum that they encountered initially, as a test of recognition. Study results showed that the round format aided recall and recognition better than the precise format. While this finding is encouraging, it is unclear whether the benefits seen from rounding in this setting (abstract math problems where one is instructed to memorize numbers) transfer to more realistic settings (e.g., news stories that contain a host of numbers, without any instructions to explicitly memorize them). Likewise, it is unclear how these effects extend to estimation problems, where one not only recalls certain numbers but also manipulates them to perform a calculation.

Although round numbers are preferred in common situations due to people's fluency with them [26], they tend to be avoided in unfamiliar scenarios – a phenomenon called the *illusion of lie effect* [17]. This effect is due to the perception that the likelihood of being able to precisely measure an unfamiliar quantity with round numbers, which are already ubiquitous in daily life, is extremely low. For instance, when an expedition team measured Mt Everest's height to be at exactly 29,000 ft, they opted to report it as 29,002 ft to avoid doubt that their measurement was rounded [1]. Likewise, prior studies have shown that people avoided using round numbers when estimating seemingly random quantities, such as the winning lottery ticket [47] or the worldwide panda population [17]. While this phenomenon is not the focus of our research, it may potentially influence the interpretation of our results, as we will discuss later.

2.2 Number rounding conventions

Conventions in number rounding often revolve around the notion of significant figures, which denote the digits that are reliable and necessary to indicate a particular quantity. While several factors come into play when deciding whether a digit is significant (e.g., the measurement resolution, the digit's location, whether it is zero), the general consensus is that one should not report more significant digits than warranted, i.e., more precision isn't always desirable.

For instance, several publication guidelines, such as the American Psychological Association [5], Academy of Management [36], and European Association of Science Editors [37], recommend rounding results to two or three significant digits. Statisticians have also argued that reporting more digits than what the data support can result in the *precision fallacy* that leads to a tendency to over-interpret one's findings [9, 13, 21, 35].

At the same time, the above guidelines and heuristics apply mostly to numbers with decimal digits in scientific writing. In more casual contexts (e.g., news reports), care should be taken to maintain a balance between readability and precision. The handbook for journalists by Livingston and Voakes [32] introduced some rough rounding suggestions – such as preserving at most one decimal value for numbers with decimals, and two or three leading significant digits for integers – while acknowledging that "there are no hard and fast rules." On the other hand, Ehrenberg [18] provided a stronger argument for rounding to two effective digits, based on the notion that only two-digit numbers can be reliably retained in memory under cognitive interruptions [43, p.40], and that precision beyond three effective digits rarely matters; however, the author offered no empirical evidence for their guidelines.

Our research seeks to address this gap by exploring the trade-off between readability and precision through empirical studies. While it may seem obvious that round numbers should be easier to recall than precise ones, here our primary contribution is investigating whether presenting less precise numbers can, perhaps counter-intuitively, lead to more precise answers.

2.3 Tools that assist with numerical comprehension

While not related to number rounding specifically, many tools developed by the HCI community have addressed the more general topic of enhancing numerical comprehension for laypeople. As an example, Hullman and colleagues [24] developed re-expression tools that implement strategies to help people better understand physical measurements that appear in news stories and data reports (e.g., 7,700 pounds is "about the weight of a car"). The two primary components of their tools include i) a database of familiar objects that can be employed in the re-expressions, and ii) an objective function to select the most effective re-expression objects. The tools were deployed as web-based applications that present the re-expressions in a pop-up view when the user clicks on a measurement within an article page. Evaluation studies showed that the tools were considered helpful based on survey ratings, although viewing the re-expressions did not reliably lead to more accurate measurement estimations, which may be due to the effort needed to interpret such re-expressions.

In another line of work, Riederer, Hofman, and Goldstein [41] used *perspective sentences* that employ ratios, ranks, or unit changes to contextualize unfamiliar numbers, with a focus on the area and population of geographic entities (e.g., re-expressing 251,827 square miles as "about the size of Texas"). This work extended prior perspective generation tools [6, 12, 28] by examining people's comprehension of, rather than preference for, different types of perspectives. Through evaluation of several perspective generation policies (e.g., one policy focuses on minimizing the objective estimation error, while the other on personalizing to the user's background), the authors found that seeing perspectives from any policy led to significantly lower estimation error, in both the short term and long term, than not seeing any perspectives. The perspective models from [41] have been deployed in platforms including the Bing search engine, where it improves instant answers to queries with numerical results, as well as Microsoft PowerPoint and Word, where it provides authors with suggestions that help their readers better understand numbers in documents.

Looking over the literature, we see that improving numerical comprehension has been a goal of many authors in HCI. One means of aiding people is via technologies that make numbers easier to visualize or reason about analogically through external stimuli. In this work, we instead explore making changes to the numbers themselves. Unlike interventions that improve cognition by presenting values in ways that are mathematically equivalent but psychologically different [11, 20], the act of rounding introduces error. Compared to seeing precise values, the error introduced by rounding could make it more difficult to approximately recall and estimate precise values. In what follows, we shall check whether this turns out to be the case.

3 EXPERIMENT 1: EFFECTS OF SEEING ROUNDED NUMBERS ON RECALL AND ESTIMATION

In this pre-registered experiment[1], we are interested in the effects of showing round versus precise numbers on people's ability to approximately recall numbers and to make estimates based on them. Specifically, we asked people to read news articles containing numbers and manipulated whether they saw the original, precise versions (with no trailing zeros) or rounded versions (rounded to one leading digit, which we will define shortly) of those numbers. We refer to the former as the Precise condition and the latter as the Round condition. Our research questions are as follows:

RQ1: *When people attempt to recall rounded versions of precise numbers, are their responses more likely to be near the precise numbers, compared to people who attempt to recall the precise numbers themselves?*

RQ2: *Does seeing rounded numbers help people make approximately correct estimates, compared to seeing the precise values?*

In examining these questions, we are asking if one can *increase accuracy* on recall and estimation tasks by *decreasing the precision* of the numbers that people are shown. Because there are multiple ways to define the precision of a number, a small clarification in terminology is merited. In a number, we denote the *leading digits* to be those that are not trailing zeros. For example, 1,200 would have two leading digits, while 1,030 would have three. We chose leading digits as our principle metric for precision over related concepts because we are interested in how the appearance of numbers affects cognition. One can easily perceive the number of leading digits in a number, even when the number of digits it has been rounded to is ambiguous. For example, 5,000 clearly has one leading digit (5), but could be the result of rounding a precise number to the nearest 1,000, 100, or 10, which cannot be disambiguated without

[1] https://aspredicted.org/ng7qa.pdf

additional context. We also note that leading digits are not the same as significant digits, whose designation takes into account measurement resolution and other factors outside the focus of our studies.

3.1 Study Design and Participants

We recruited 1,300 participants on Amazon's Mechanical Turk platform [34] to obtain 95% power in detecting a 10 percentage point difference with a 5% false positive rate, based on results of prior pilot studies. Participants were required to live in the U.S., have completed at least 100 Mechanical Turk tasks on the platform, and have a high approval rate on previous tasks (≥ 99%). In addition, they must not have completed any prior pilot study on the same subject. They received $1.50 compensation upon completion of the experiment.

Participants were randomly assigned to either the Precise condition or the Round condition. Before seeing any stimuli, participants were told that they would be shown a few news articles and asked questions about them; no special mention was made of the fact that they contained numbers, and no indication of the type of questions they would be asked was provided[2]. After confirming that they understood these instructions, participants were shown three news article snippets, where each snippet contained two focal numbers. The articles in the two conditions differed only in how the focal numbers were represented (Figure 1). In the Precise condition, the focal numbers were shown in precise format, with no trailing zeros (e.g., $41,282$). In the Round condition, the same focal numbers were rounded to one leading digit (e.g., $40,000$). In total there were three sentences containing six focal numbers; each sentence belonged to a separate article snippet and was highlighted in the same manner as in Figure 1. We have used this highlighting strategy in past work [6] and found it to be helpful in focusing people's attention; without the highlight we would expect lower performance overall, but would not expect a differential effect between the randomly-assigned round and precise conditions.

The highlighted sentences are listed as follows, with values that participants in the Precise condition saw inline and values that participants in the round condition saw in parentheses:

(1) Among the admitted students were 28,752 (30,000) transfer students who were offered spots at U.C campuses, out of 41,282 (40,000) applicants [15].
(2) By 2000, the number of Honduran immigrants in the United States, mostly without proper visas, was 282,852 (300,000) and it now stands at 487,745 (500,000) according to a Migration Policy Institute report [4].
(3) Of the 40,108 (40,000) square feet the Frick wants to add, only 3,792 (4,000) of it would be for showing art [30].

The snippets used in our study were sampled from a wide range of articles published in the New York Times[3] and selected based on two criteria: each should have two numbers in a focal sentence, and the three articles should cover different number magnitudes. They were presented in a fixed order, as listed above, so as to maintain a consistent time gap between the reading task and the subsequent cognitive tasks for each article.

Next, participants were asked to recall the six focal numbers from the snippets (Figure 2). We coded each response as *approximately correct* if it was within 10% of the *precise* version of the number, as per our pre-registration plan. For example, if the precise number was 41,282, any answer within 10% of this number (from 37,153 to 45,410) would be coded as approximately correct. This criterion applied to both conditions, putting participants in the Round condition at a disadvantage as they never saw 41,282, but only its rounded version, 40,000. We have adjusted the values of the focal numbers in the above articles so that all the rounded numbers were within 10% of their precise counterparts; in this way, participants in the Round condition could still be approximately correct if their recall responses were close to the round numbers they saw.

For each of the three snippets, participants were then asked to estimate a new number based on their memory of the focal numbers they read (Figure 3). Our goal was to see which number format would be easier to work with mentally, i.e., whether it is easier to perform calculations with numbers recalled from the round or precise formats. Similar to the recall task, we coded each response as approximately correct if it was within 10% of the answer one would calculate from the *precise* version of the focal numbers. For example, the task in Figure 3 involved estimating 28,752 as a percentage of 41,282 for those in the Precise condition. Answers within 10% of the ground truth value (69.65%) were coded as approximately correct. This criterion applied to both conditions, despite participants in the Round condition never having seen 28,752 and 41,282 and having seen 30,000 and 40,000 instead. As with the recall task, this criterion intentionally puts the Round condition at a disadvantage to provide for a more conservative test. Our adjusted focal numbers also ensured that the estimation values derived from the round numbers were all within 10% of those derived from the precise numbers.

To deter participants from cheating (e.g., recording the given numbers, searching for them online) or producing undesirable responses, we pre-registered and implemented two exclusion mechanisms. The first was an attention check question where participants read the experiment instructions and checked the options that correctly described what they should and should not do (Figure 4). Out of the five given options, only four should have been checked. Participants were given up to two chances to check the correct set of options; those who failed both times were excluded from our analysis. The second mechanism was a recall question that involved a large, precise number – the total population of California (39,613,493 people), shown at the end of one of the three snippets (see Figure 1 where this number was presented in precise format in both conditions). We excluded participants who recalled this number exactly as it was presented, on the grounds that they either cheated or had exceedingly good recall abilities (i.e., were outliers). In total, we excluded 204 participants based on the first mechanism, 44 based on the second mechanism, and an additional 5 who entered invalid study completion codes; however, these participants were still compensated at the normal rate for their work. Our final sample consists of 1047 participants, with 547 in the Precise condition and 500 in the Round condition.

[2]The full set of stimuli shown to participants in this experiment is included in the supplementary file `experiment1_content.pdf`.
[3]https://www.nytimes.com/

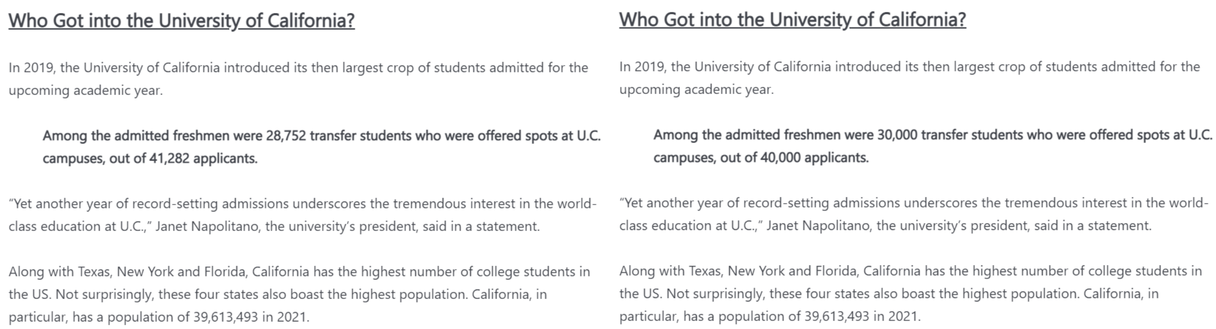

Figure 1: A news article snippet [15] in the reading task for the Precise condition (left) and Round condition (right). The focal numbers are included in the bolded sentences.

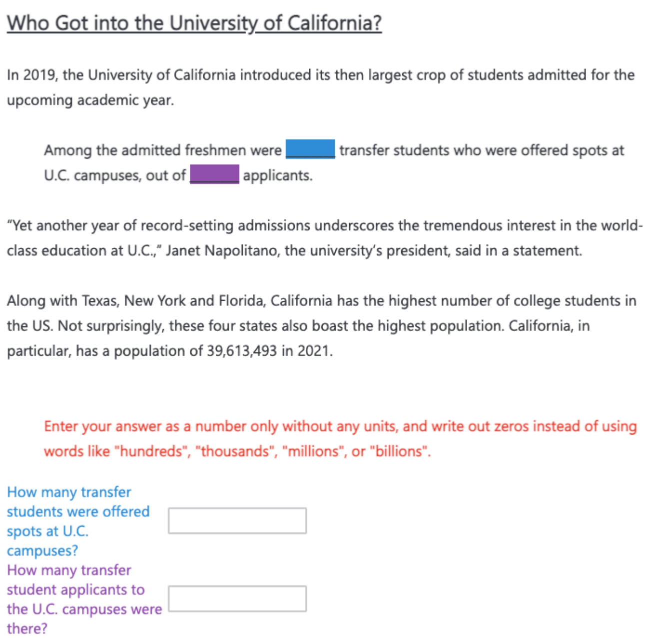

Figure 2: A recall task based on the news article snippet in Figure 1.

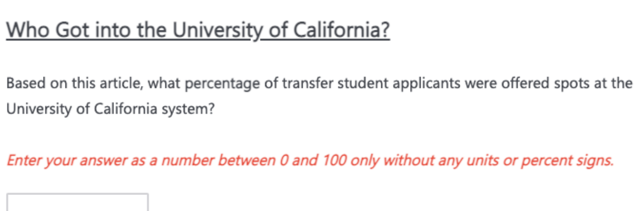

Figure 3: An estimation task based on the news article snippet in Figure 1.

This is a psychology experiment. We are interested in how people think. You are not being paid based on the correctness of your answers. You are supposed to take your best guesses when answering questions. Please do not look up information on the internet. Please do all work in your head. Please do not write anything down.

Based on the paragraph above, please check all statements that apply.

- ☐ This study is a psychology experiment and I'm not being paid based on the correctness of my answers.
- ☐ I'm supposed to take my best guesses.
- ☐ I'm not supposed to look up answers on the internet.
- ☐ I'm supposed to type random answers, not my best guesses.
- ☐ I'm supposed to work from information in my head and not record or write down any information I see.

Figure 4: The attention check question, with one incorrect option "I'm supposed to type random answers, not my best guess."

3.2 Results

Following our pre-registration, we first report the results of our main research questions, which involve the recall and estimation accuracy of participants in each condition. Then we present a number of secondary analyses to better understand behavioral differences between those who were shown either precise or round numbers.

3.2.1 When people attempt to recall rounded versions of precise numbers, are their responses more likely to be near the precise numbers, compared to people who attempt to recall the precise numbers themselves? Looking at the proportion of approximately correct answers (those within 10% of the precise value) across all six recall tasks (Figure 5, left), we see a large and statistically significant benefit from displaying round numbers: fewer than 50% of all responses in the Precise condition were approximately correct, compared to almost 70% in the Round condition ($\hat{p}_{\text{precise}} = 0.49$ vs. $\hat{p}_{\text{round}} = 0.68$; $Z = 265.36, p < .001$, 95% CI $(-0.21, -0.16)$ in a two-sided proportion test). Within each individual recall task (one task for each focal number), we observed a similar pattern, where participants in the Precise condition were about 20 percentage points less likely to provide an approximately correct response, compared to those in the Round condition (Figure 5, right).

Secondary analyses. Our primary research question concerned how often people are approximately correct when presented with precise or rounded numbers. While computing the approximately correct proportions allows for a straightforward comparison between conditions, an analysis of the full error distribution can further reveal how often errors of various magnitudes are made in each condition. Figure 6 shows, for the Precise and Round condition, the distribution of recall error percentages computed relative to the precise values. For example, a response of 1,500 compared to a precise value of 1,000 would constitute a 50% error. The leftmost bars in each panel correspond to errors of less than 10% and thus report the same "approximately correct" proportions shown earlier in the left panel of Figure 5. At the same time, looking at the bars that correspond to other error margins, we observed lower values for nine out of ten bars in the Round condition. In other words, people presented with round numbers were less likely to make errors at nearly all of the margins shown, not just at the 10% margin. Even for the most extreme case (error margins of at least 100%, corresponding to the rightmost bars), participants in the Round condition were 50% less likely to fall into this case than those in the Precise condition. Interestingly, we observed a slight increase in the frequency of recall errors at the 90-100% margin, compared to the other error margins beyond 10%. Upon closer inspection of the responses in this error category, we found that they were mostly correct in the first digit but underestimated the target number length (e.g., recalling 40,108 as 4,000). Likewise, responses at an error margin above 100% were mostly due to their overestimation of the target number length (e.g., recalling 40,108 as 400,000).

In another secondary analysis, we considered the time taken across all the recall tasks in each condition. A Wilcoxon signed rank test showed a small but statistically significant effect on the total recall time, $W = 157,500, p < .001$, where those in the Precise condition (median = 70 seconds) took more time than those in the Round condition (median = 62 seconds).

3.2.2 Does seeing rounded numbers help people make approximately correct estimates, compared to seeing the precise values? We computed the proportion of approximately correct answers across all estimation items and participants in each condition. A two-sided proportion test showed a large and statistically significant effect on the approximately correct proportion, $Z = 339.43, p < .001, 95\%$ CI $(-0.36, -0.29)$, with the Round condition leading to more than twice the rate of approximately correct estimates compared to the Precise condition ($\hat{p}_{\text{round}} = 0.60$ vs. $\hat{p}_{\text{precise}} = 0.27$). A similar pattern was observed within each individual estimation task, where participants in the Round condition were about 15-45 percentage points more likely to make an approximately correct estimation than those in the Precise condition (Figure 7).

Secondary analyses. To better understand the difference in the quality of estimates across conditions, we looked at the consistency between the values people recalled and the estimates they made. In particular, we compared the *recall-estimate consistency* (i.e., whether the participant's estimation response could be reconstructed from their recalled responses) between the two conditions. As an example, based on the article snippet in Figure 1, our recall tasks involved recalling the number of transfer applicants (41,282) and number of admitted students (28,752) to the University of California system, while the estimation task involved estimating the percentage of accepted transfer students (69.65%). In this case, we could compare a participant's *reported estimate* (i.e., their response to the estimation task) to their *implied estimate* (i.e., the calculation result based on their responses to the recall tasks). If these values are equal, a participant is considered consistent in their recall and estimation; that is, they did not make any calculation errors while using the recalled numbers to make their estimate. As shown in Figure 8, participants in the Precise condition were about 40-55 percentage points less likely to be consistent than those in the Round condition. In other words, we found that round numbers led to higher accuracy in both recall *and* calculation, which then contributed to better estimation performance by participants in the Round condition. The same pattern holds if we look at approximate instead of exact agreement between recalled and estimated values (i.e., whether the implied estimate is within 10% of the reported estimate)[4].

Additionally, we considered the time taken across all the estimation tasks in each condition. A Wilcoxon signed rank test showed a small but statistically significant effect on the total estimation time, $W = 146,407, p = .048$, where those in the Precise condition (median = 47 seconds) took more time than those in the Round condition (median = 44 seconds).

While this experiment reveals that rounding numbers has clear benefits on readers' numerical comprehension, it leaves open the question about readers' preferences for seeing round versus precise numbers. We designed our second and third studies to survey people on their preferences for when rounding is appropriate and what degree of rounding is preferred across a range of magnitudes and contexts.

4 SURVEY 1: PREFERENCE FOR ROUNDING DEGREE

This survey consisted of a series of tasks where we showed people hypothetical news quotes that contained a precise number and asked them to rate several rounded versions of the number, ranging from preserving full precision to fully rounding it (i.e., rounding to one leading digit). Across the tasks, we varied the context in which the number appeared (referring to either money, population, or distance) as well as the length (magnitude) of the number, ranging from thousands to billions. Our research questions are as follows.

RQ3: *What degree of rounding is considered acceptable for numbers of different magnitudes and in different contexts?*

RQ4: *What is the smallest number of leading digits that is considered acceptable for numbers of different magnitudes?*

4.1 Study Design and Participants

We recruited 80 workers on Amazon's Mechanical Turk platform [34] to participate in the study, in exchange for $2 compensation

[4]See the graph `approx_recall_estimate_consistent.png` in the supplementary materials.

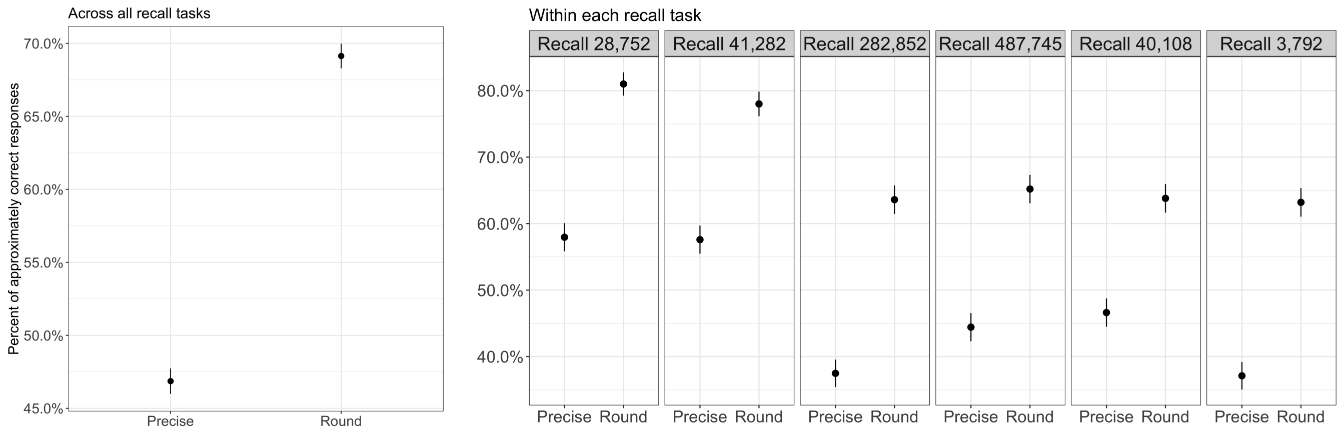

Figure 5: Recall accuracy relative to precise values, across all tasks (left) and in each individual task (right). Error bars denote one standard error.

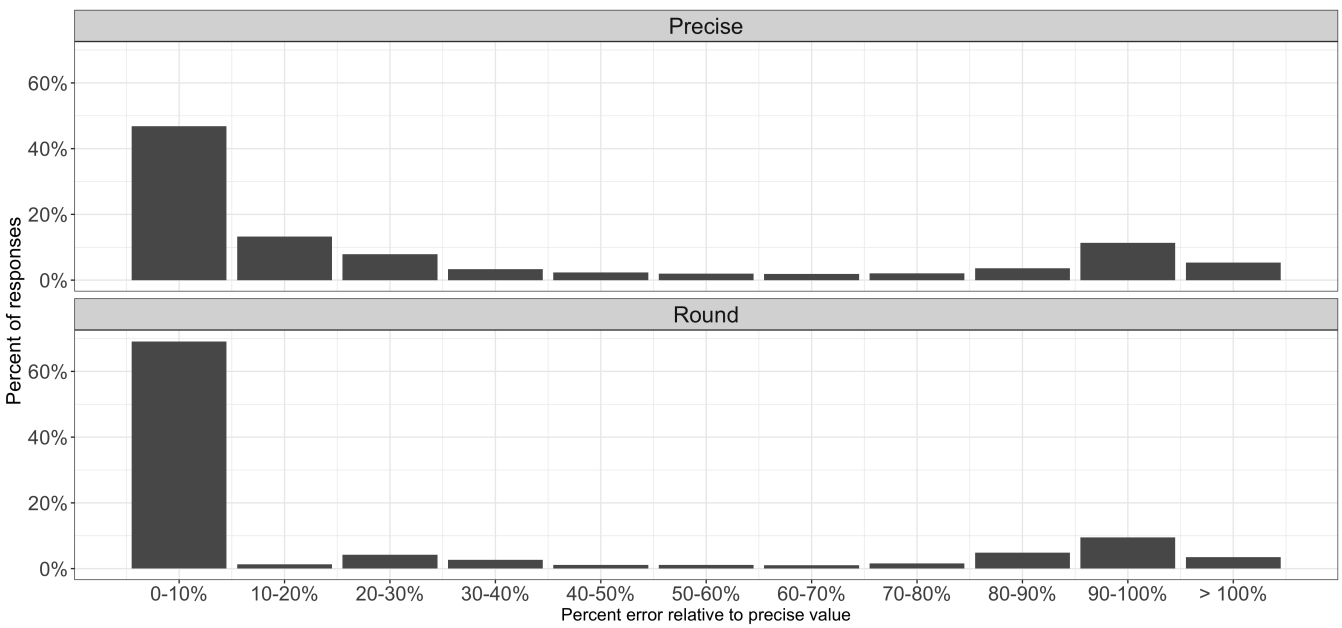

Figure 6: Distribution of percent error in recalling precise values by condition. Bins labeled with intervals are left inclusive and right exclusive.

upon completion. Participants were required to live in the U.S., be classified as a Master Worker on the platform, have completed at least 100 Mechanical Turk tasks, and have high approval rate on previous tasks (≥ 99%).

Each participant completed a survey with 15 tasks. The interface for each task is as depicted in Figure 9, where participants were asked to provide rounding suggestions for a given number to a hypothetical newspaper editor, by rating each rounding option as either "Too precise," "Looks good" or "Too round." We display short explanations below each option (e.g., "Too precise" is explained as "few readers would want this much precision") to clarify the rating criteria. We also include a reminder that the participant’s selections should be *sensible*, i.e., no answer should be to the left of the answer above it. For example, if they rated rounding to four leading digits as "Looks good," then any option that corresponds to more rounding should only be rated as either "Looks good" or "Too round," not "Too precise." This constraint was not enforced in the rating interface, but was used afterwards to identify low-quality responses and eliminate participants, as we discuss below.

For each participant we generate 15 artificial news quotes based on the following procedure. First, each quote comes from one of three templates, as listed in Table 1. We use each template five times,

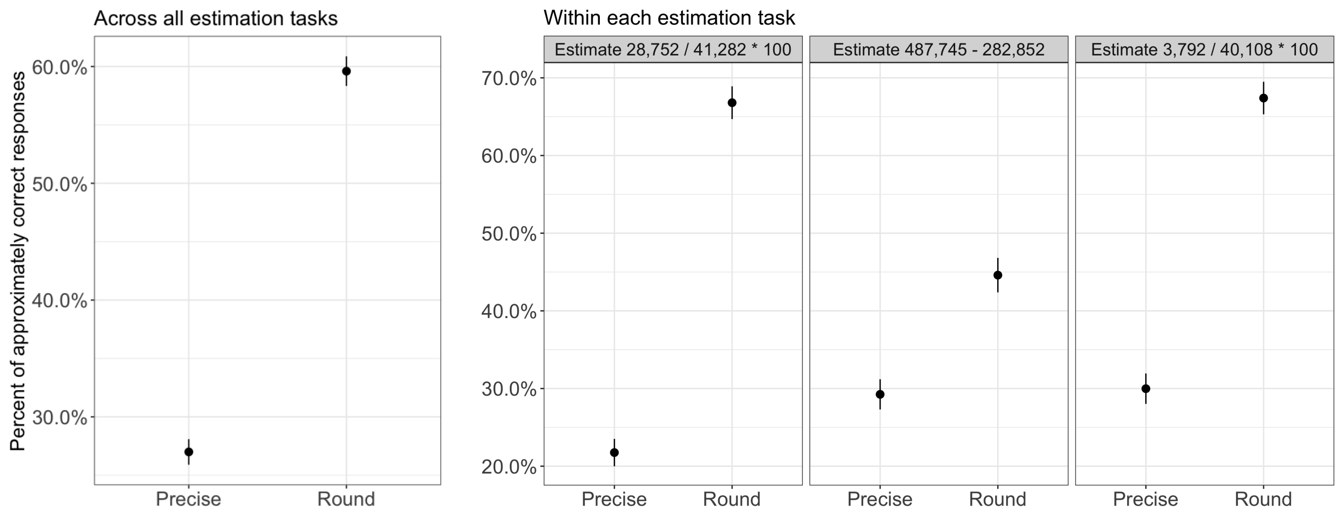

Figure 7: Estimation accuracy across all tasks (left) and in each individual task (right). Error bars denote one standard error.

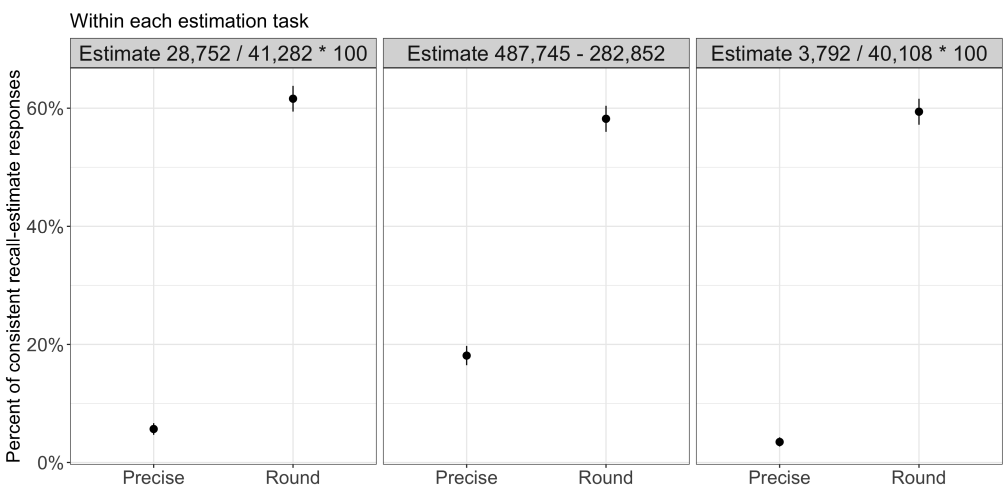

Figure 8: Percentage of responses that are consistent with recalled values. Error bars denote one standard error.

so that each participant sees five quotes about money, five about people, and five about distance. To fill in the number in each quote, we first randomly select the two leading digits for each number, out of a total of 90 possible combinations of two leading digits – from 10 to 99. Then we assign each of these an order of magnitude, taking three samples at each of five different levels (4 digits, 6 digits, 7 digits, 8 digits, or 10 digits). We selected these thresholds to cover a wide range of number magnitudes, with the assumption that the missing thresholds (5 digits and 9 digits) can be interpolated from our results. Once the two leading digits and order of magnitude for each number have been determined, we sample the remaining digits uniformly at random between 1 and 9, thus forming a precise value for each number shown in the quote. We display these quotes in a randomized order to participants as Tasks 1 through 15. Within each Task, all possible rounding options for the given number are then generated and ordered from the lowest to the highest degree of rounding, as seen in Figure 9. If the given number is at least a million, we include the appropriate modifier term in each rounding option (e.g., "million," "billion"), to be consistent with how large numbers are typically reported in the news [40].

To ensure quality responses, we excluded eight participants who had fewer than 80% (12 out of 15) sensible task responses; however, these participants were still compensated at the normal rate for their work. Our final sample includes 72 participants, who provided 1062 sensible task responses in total.

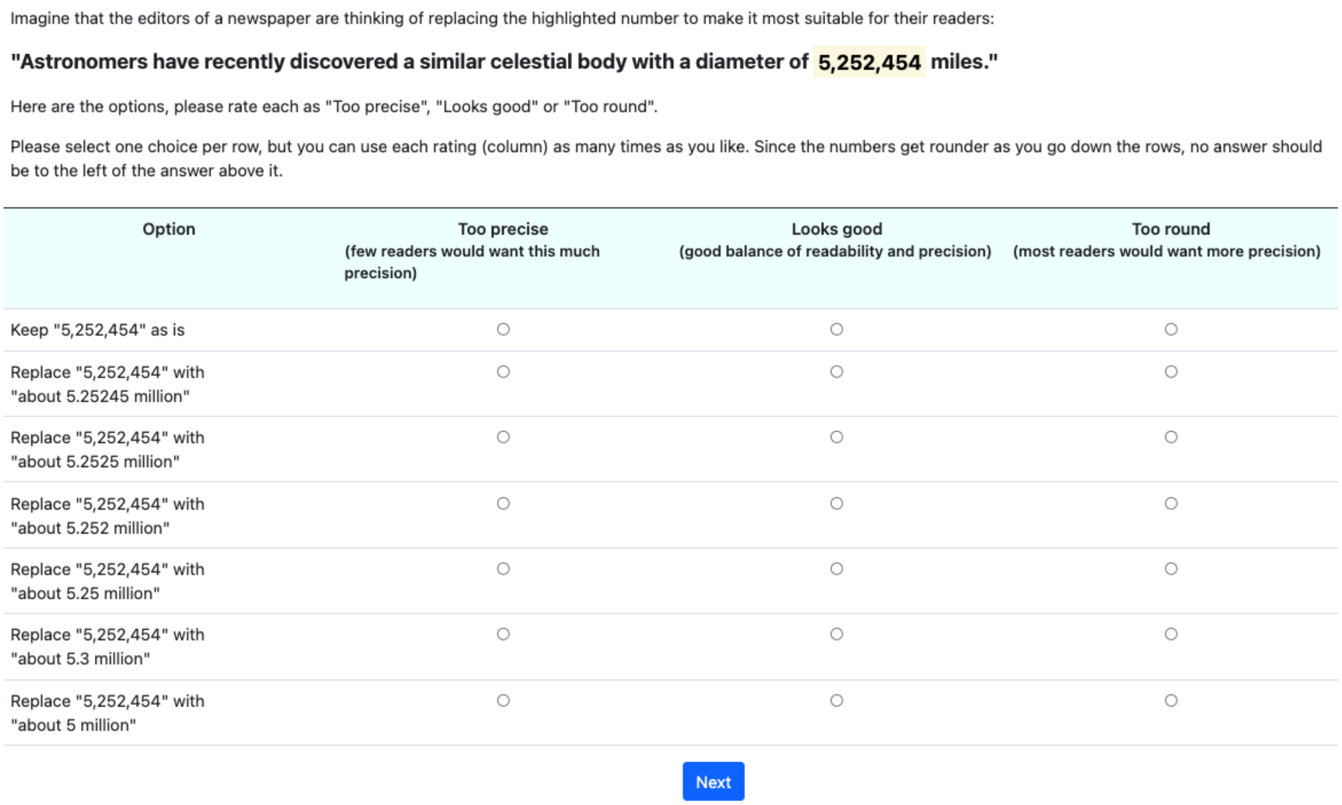

Figure 9: Screenshot of a task in the rating interface of Survey 1.

Table 1: The template sentences used in the survey and their corresponding contexts. The dashed portion indicates where a randomly generated number is to be filled in.

Context	Template sentence
Money	In a recent report, the hospital group disclosed that it spent $_____ on healthcare equipment.
People	DNA evidence suggests that throughout history, _____ people have been affected by the disease.
Size	Astronomers have recently discovered a similar celestial body with a diameter of _____ miles.

4.2 Results

Here we use the term *full number length* to refer to the length of the given number in a task (which is either 4, 6, 7, 8, or 10 digits), and the term *acceptable* to refer to the rounding options rated as "Looks good."

4.2.1 What degree of rounding is considered acceptable for numbers of different magnitudes and in different contexts? Figure 10 depicts the percentage of participants who rated each rounding option (represented by their leading digit count) as acceptable. We observed a consistent pattern across different number magnitudes (i.e., different line graphs), where about 80% of participants considered preserving two or three leading digits as acceptable. In addition, about 60% of participants rated four leading digits as acceptable when the full number length was four (in other words, they preferred that four-digit numbers are not rounded at all), but this percentage dropped sharply for longer numbers. Finally, the line graphs look very similar across the three contexts, suggesting that these contexts had little effect on people’s rounding preferences.

4.2.2 What is the smallest number of leading digits that is considered acceptable for numbers of different magnitudes? For each participant and in each task, we identified the highest degree of rounding rated as "Looks good" by them and recorded the *minimum acceptable leading digit count*[5] of this rounding option. For example, in Figure 9, if a participant rates the second-to-last option "about 5.3 million" as "Looks good" and the last option "about 5 million" as "Too round," then "5.3 million" represents the highest degree of rounding acceptable to that participant, which corresponds to a minimum acceptable leading digit count of two. Based on this definition, we plotted the average minimum acceptable leading digit count at each full number length in Figure 11. Here we observed that people were comfortable with rounding to two or three leading digits, even as the full number lengths varied from four to ten. In addition, it appears that people may have different preferences for numbers when they were expressed as digits only (the leftmost two points on the graph, representing four- and six-digit numbers) versus when a modifier like "million" and "billion" was used (the

[5] As a reminder, we denote the *leading digits* to be those that are not trailing zeros.

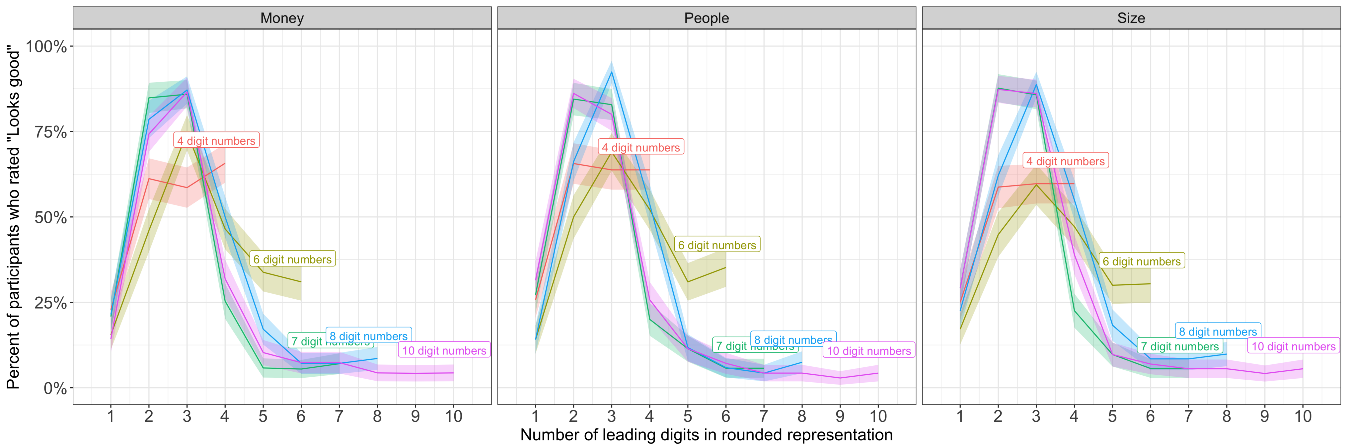

Figure 10: Percentages of "Looks good" ratings across different full number lengths and contexts.

rightmost three points, representing seven-, eight-, and ten-digit numbers). When only digits were present, we saw that the minimum number of acceptable leading digits increased with the magnitude of the number (for four-digit numbers, at least two leading digits were deemed acceptable, whereas for six-digit numbers, at least three leading digits were deemed acceptable, on average). When "million" or "billion" was used, however, at least two leading digits were found to be acceptable across a range of full number lengths.

Based on the above results, we were interested in why people largely converged on rating two or three leading digits as the highest acceptable degree of rounding, regardless of the given number magnitudes. One hypothesis is that for long numbers (e.g., 5,252,454) people considered the rounded options such as "5.25 million" and "5.3 million" acceptable because they treated the thousandth and subsequent digits as insignificant. That is, they are used to seeing only one or two digits after the decimal point. The follow-up question, then, is whether this perception persists if we fully write out the rounding options, without any decimal point or modifier term (e.g., "5,250,000" instead of "5.25 million"). We conducted a second survey to examine this question.

5 SURVEY 2: PREFERENCE FOR ROUNDING DEGREE (WITHOUT DECIMALS OR MODIFIERS)

The setting for this survey was mostly similar to that for Survey 1, where we asked participants to rate different rounding options of a given number presented in a template sentence. A notable difference was that the decimal points and modifier terms ("million" and "billion") were no longer present in the rounding options. We examined the same research questions as we did previously:

RQ3': *What degree of rounding is considered acceptable for numbers of different magnitudes and in different contexts?*
RQ4': *What is the smallest number of leading digits that is considered acceptable for numbers of different magnitudes?*

5.1 Study Design and Participants

We recruited 234 workers on Amazon's Mechanical Turk platform [34] with similar qualifications as in Survey 1: participants needed to be Masters Workers who live in the U.S., have completed at least 100 Mechanical Turk tasks, and have high approval rate on previous tasks (≥ 99%). In addition, they must not have completed Survey 1.

Each participant completed a survey with 15 tasks in the same manner as in Survey 1, for a compensation of $2. The tasks with four- or six-digit numbers were identical to those in the previous survey, while the tasks with longer numbers include rounding options that are fully expressed in digits with trailing zeros. For example, the option "about 5.2525 million" in Survey 1 (Figure 9) is translated to "about 5,252,500" in Survey 2. We used a larger sample of participants in Survey 2 and preserved the questions about four- and six-digit numbers to better highlight both the similarities and differences in findings from the two surveys.

Following our quality check, we excluded 19 participants who had fewer than 80% (12 out of 15) sensible task responses; however, these participants were still compensated at the normal rate for their work. Our final sample includes 215 participants, who provided 3177 sensible task responses in total.

5.2 Results

5.2.1 What degree of rounding is considered acceptable for numbers of different magnitudes and in different contexts? Figure 12 depicts the percentage of participants who rated each rounding option (represented by its leading digit count) as acceptable. In contrast to Figure 10 from Survey 1, there was no longer a convergence in the acceptable ratings at two and three leading digits; instead, as the full number lengths increased, people tended to only rate options with more precision as acceptable. Across the three contexts, the line graphs remain similar, indicating that the contexts (money, people, or size) again had little effect here.

5.2.2 What is the smallest number of leading digits that is considered acceptable for numbers of different magnitudes? Figure 13 depicts the distribution of the minimum acceptable leading digit count at each full number length. We observed that, in this survey, the

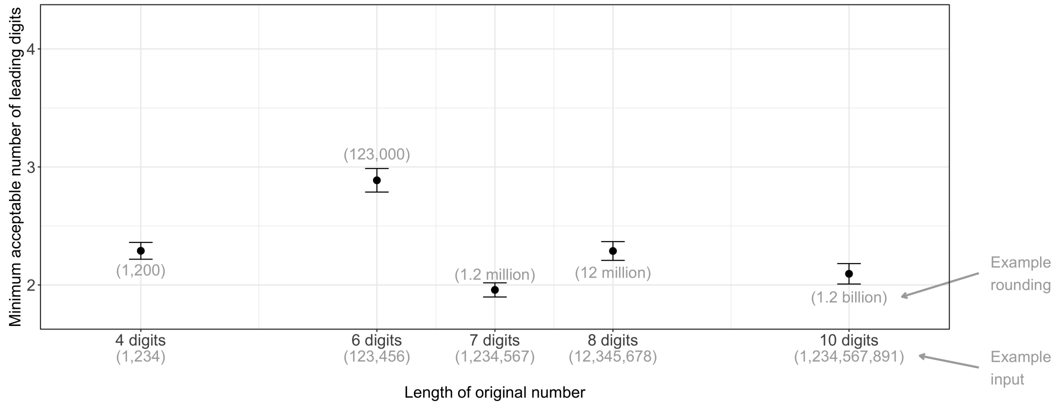

Figure 11: The average minimum acceptable leading digit count across the five full number lengths. Grey annotations show example input numbers (on the *x*-axis) and how they would be rounded based on the average minimum acceptable leading digit count (above or below each point on the plot). Error bars denote the magnitude of one standard error.

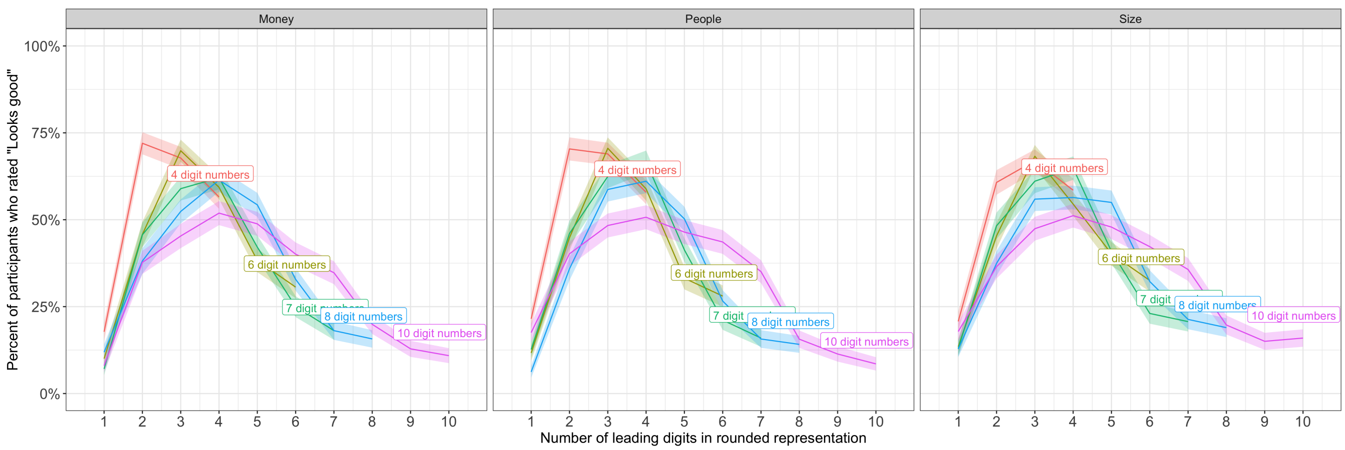

Figure 12: Percentages of "Looks good" ratings across different full number lengths and contexts.

minimum acceptable precision level increased linearly with the full number length. In other words, with larger given numbers, more leading digits were preferred for the highest acceptable rounding options. From comparing the two distribution groups in Figure 11 and Figure 13, we noted that this linear pattern was not present in Survey 1, so the digit-based representation in Survey 2 appears to have an impact on people's preferences for the minimum precision level. The implications of these format-dependent preferences are considerable. For ten-digit numbers, people preferred rounding to the nearest one million in one format (with digits only) and to the nearest 100 million in the other (with modifier terms). Using acres as an example, this is akin to the difference between measuring in units the size of Rhode Island and units the size of California. Upon inspection of the full distribution of ratings[6], we observed that responses were centrally distributed in each group of full number length. For example, when the given number had seven digits, roughly one third of participants preferred a minimum of two leading digits, while a quarter preferred a minimum of three and 15% preferred a minimum of four.

Finally, we constructed mixed effect models to capture people's preferences for the minimum acceptable leading digit count. As the rating data points were not independent (each participant provided several ratings), we used the participants' anonymized IDs as a random effect. In this setting, we examined two models: a baseline model M_1 that only uses the random effect, and a full model M_2 that also incorporates features of the rating task, namely the full number length and context factor:

$$\begin{aligned} M_1 :&\quad \text{Minimum Acceptable Leading Digits} \\ &\quad \sim (1 \mid \text{Participant ID}), \\ M_2 :&\quad \text{Minimum Acceptable Leading Digits} \\ &\quad \sim \text{Full Length} + \text{Context} + (1 \mid \text{Participant ID}). \end{aligned}$$

[6]See the graph `survey2_full_rating_distribution.png` in the supplementary materials.

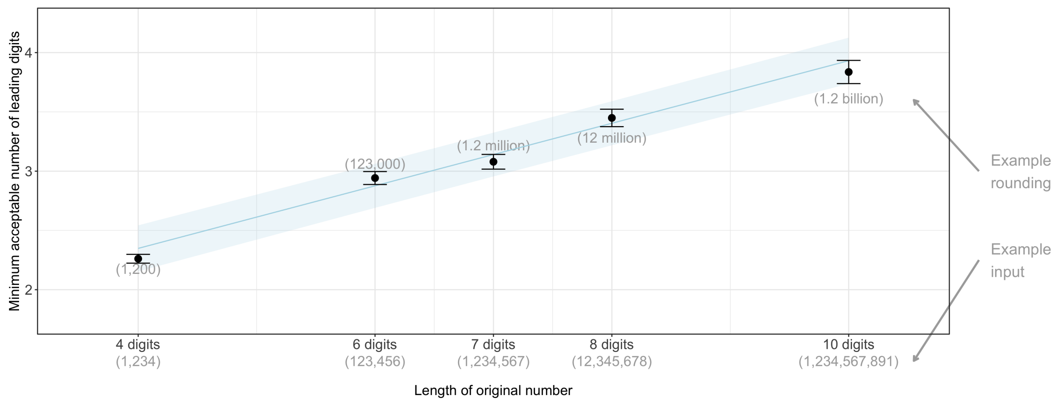

Figure 13: The mean and standard error of the minimum acceptable leading digit count across the five full number lengths. Grey annotations show example input numbers (on the x-axis) and what they would be rounded to at the average minimum acceptable number of leading digits (above or below each point on the plot). The blue line and ribbon show a model fit to all of the responses (model M_2 described in the text), averaging over fixed effects (context) and random effects (participants).

We evaluated these models using two metrics: the mean squared error (MSE) and R^2, computed by the `modelr` [50] and `MuMIn` [7] packages in R. In addition, the p-values for the fixed effects were estimated with t-tests using the Satterthwaite approximations for degrees of freedom [42]. The baseline model M_1 had an R^2 of 50% and MSE of 1.52, while the full model M_2 had an R^2 of 59% and MSE of 1.24. The estimated fixed effects for M_2 are included in Table 2, which shows that the full number length had a positive, statistically, and practically significant effect even within people, while the context was not a significant factor. The variance of the random effect Participant ID was 1.63. We therefore observed that, while individual preferences for rounding account for a good amount of the variance in responses, people generally prefer to see larger numbers rounded to more leading digits when the numbers are written out as digits only (without modifiers like "millions" or "billions").

Table 2: Estimated model fixed effects for model M_2. (*) indicates significance at the $\alpha = .001$ level.

Effect	Estimate	Std Error	t	95% CI
Intercept	1.29 (*)	0.12	10.93	$(1.06, 1.52)$
Full Length	0.26 (*)	0.01	25.80	$(0.24, 0.28)$
Context: People	-0.06	0.05	-1.20	$(-0.16, 0.04)$
Context: Size	0.02	0.05	0.39	$(-0.08, 0.12)$

6 DISCUSSION & CONCLUSION

Our work tested for and identified beneficial effects of presenting people with round numbers instead of precise ones. Readers of round numbers were considerably better at approximate recall and less likely to make large errors in what they recalled. They were likewise better at putting what they recalled to work: they performed more accurate calculations with the recalled numbers and arrived at better estimations. We also conducted two survey studies that uncovered clear and format-dependent preferences for how numbers should be rounded. The effect of the number format on preferences was considerable, with people preferring up to 100 times more precision in one format (having digits only) than in another (having modifier terms). In this section, we further discuss the findings of our studies and how they can be leveraged to enhance consumption of digital content.

6.1 Cognitive effects of round numbers

In our first study, we asked participants to read six focal numbers (either with full precision or rounded to one leading digit) embedded in three news article snippets, then to recall and perform estimations based on these numbers. Our evaluation criterion was whether the participant's response was within 10% of the *precise* numbers. We note that this criterion put participants in the Round condition at a disadvantage because they never got to see the precise numbers that they were evaluated against. Nevertheless, there was a large advantage in both recall and estimation accuracy for those who saw round numbers. In other words, we observed a trade-off, by which presenting round (as opposed to precise) numbers results in some initial *precision* loss due to rounding, but leads to a much larger *accuracy* gain during subsequent recall and estimation with these numbers. Additionally, we found that those who saw round numbers took slightly less time on average to perform recall and estimation than their counterparts, which is also indicative of a gain in cognitive *efficiency*.

Our finding is consistent with past research showing that people tend to work better with round numbers in various contexts, for example communicating clock times [45] or recalling the sums of accounting problems [33]. We have also shown that this effect is robust across different error margins used for evaluation: based on our secondary analysis (Figure 7), participants in the Round condition made fewer errors not only at the 10% margin, but across nearly all margins, including even the most extreme case (where the relative error exceeded 100%). Additionally, we have separated the effects of round numbers on recall and on calculation in our evaluation of participants' consistency. If round numbers had no effect on calculation ability, we would expect participants in both conditions to be equally good at using their recalled responses to calculate the estimation responses, that is, to have similar levels of recall-estimate consistency. Instead, our results again revealed that participants in the Round condition had higher consistency by a large margin. In other words, those who saw round numbers had more accurate estimates because they were better at both recall and calculation than those who saw precise numbers.

6.2 Preferences for degrees of rounding

While certain cognitive benefits of round numbers seem rather clear, it is not obvious that numbers should always be rounded. To elucidate this point, we conducted two survey studies that elicited preferences for when and by how much to round numbers. In particular, we asked participants to rate different rounding options for a given number in a particular context (e.g., a seven-digit number that denotes the diameter length of a celestial body), with a focus on the minimum acceptable precision level (operationalized as the count of leading digits). In Survey 1, we expressed the rounding options for large numbers using decimal values and modifier terms (e.g., "12.345 million"). In Survey 2, we replaced this format with writing out the full digits (e.g., "12,345,000"), while keeping all other aspects identical. Based on our results, while the minimum acceptable leading digit count remained at around two to three digits across all number magnitudes in Survey 1, it had a significant and positive association with the given number magnitudes in Survey 2. These two patterns differed prominently on larger numbers – for ten-digit numbers, in particular, the average minimum acceptable leading digit count went from around two in Survey 1 to around four in Survey 2 (Figure 11 and Figure 13).

Our first conjecture regarding this distinction, which was the primary motivation for conducting Survey 2, is that the decimal expressions in Survey 1 led participants to tolerate more rounding than they normally would. People are likely used to seeing only one or two digits after a decimal point and may treat any subsequent digit as insignificant. In this manner, they may also consider "1.2 million" as an acceptable approximation of "1.234567 million," even though dropping the hundredth digit (3) implies dropping by more than 30,000 in value. In Survey 2, we would instead display this same rounding option as "1,200,000," which clearly shows five digits being zeroed out. Our prediction was that this format would better convey how much precision is lost, and lead people to prefer more digits for their minimum precision level. Indeed, this interpretation is consistent with the results of Survey 2.

The use of number separators "," and "." may also play a role in the contrast between the findings of Survey 1 and 2. The comma separators can implicitly partition a number into smaller groups of digit (e.g., the commas in "1,234,567" partition it into three groups: "1", "234" and "567"), which may cause people to prefer either preserving or rounding all the digits in each group, due to aesthetic reasons. We see supporting evidence of this behavior for some number magnitudes in Figure 13, where the average minimum acceptable leading digit count was about three for six-digit numbers (e.g., "123,000") and four for ten-digit numbers (e.g., "1,234,000,000"). Conducting another follow-up experiment where the numbers are written out in full digits but without the comma separator would allow us to test this conjecture. We also note that the dot separator, which separates the integer portion from the decimal portion, did not yield similar effects. If people preferred rounding to integer when the dot separator was present, we would observe a high percentage of "Looks good" ratings at one leading digit when the given number had seven or ten digits; however, this was not the case in Figure 10.

A third explanation is that people decided the minimum acceptable leading digit count based on how much rounding error they could tolerate, rather than on the format of the number (e.g., with or without modifier terms). To guarantee at most 5% relative rounding error[7], for example, one needs to preserve just the first two leading digits of any number, regardless of its magnitude. On the other hand, to guarantee an upper bound on the absolute error[8] of, say, 50, one would need to preserve all but the rightmost two digits. If rounding error was indeed a factor of consideration, our linear model in Figure 13 and Table 2 would suggest that people's preferences lie somewhere between these two strategies of rounding to bound relative error and rounding to bound absolute error.

It is also worth pointing out that the final leading digit of a rounding option may influence its ratings, given that, other than zero, five is also a popular final digit in many contexts [27]. From a post hoc analysis of our data, we did not observe that rounding options ending in five were more likely to be rated as "Looks good" than options ending in other digits[9]; however, a follow-up study that explicitly manipulates this factor may reveal stronger patterns.

Additionally, our linear model confirmed that the context of the template sentence (money, people or size) did not significantly impact how people rated the rounding options. While we expected that people may look for more precision in some contexts (e.g., money) than in others, it is possible that such an effect, if present, would require either making the context more prominent (e.g., through an entire paragraph rather than a single template sentence) or examining a wider range of contexts. An alternative explanation is that context should encompass not just the information domain, but also other aspects of the focal numbers, such as their measurement precision, origin and use cases. For instance, on the same domain of money, an article about housing price inflation may need more precision than one that reports on a movie's opening weekend revenue, as the former number has more value in aiding the reader's decision-making. Future studies that manipulate these different dimensions could better reveal the nuances around number rounding preferences and address the lack of number rounding heuristics in current literature.

Finally, we note that our full model, which incorporated both the context and full number length, only achieved a minor improvement in R^2 and MSE, compared to a baseline model with solely the random effect Participant ID. In other words, people's individual preferences still played a significant role in their ratings of acceptable degrees of rounding.

[7]The relative rounding error is defined as $\frac{\text{original number} - \text{rounded number}}{\text{original number}}$.
[8]The absolute rounding error is defined as $|\text{original number} - \text{rounded number}|$.
[9]See the graphs `survey1_pct_looks_good_by_final_digit.png` and `survey2_pct_looks_good_by_final_digit.png` in the supplementary materials.

6.3 Limitations and Future Work

Taken together, the results of our experiment and surveys support the idea that less can be more when it comes to displaying numbers at different levels of precision. In particular, we found that readers stated preferences for suitably rounded numbers that balance readability and precision. The randomized experiment we conducted also confirmed that this preference for rounded numbers is sensible, as it led to improved recall and estimation. These effects were not only consistent across the questions we studied but also sizeable, ranging from 20 to 40 percentage point gains in approximately correct answers. Nonetheless, there are a number of limitations which influence the interpretation of our results and, at the same time, point towards avenues for future research in related topics.

In our first study, we examined only a handful of numbers in particular settings (six specific news quotes), and looked at extremes in terms of precision (no trailing zeros versus only one leading digit). To make the results more representative, future work should have the news snippets sampled from a wider range of articles and treated as a random effect. In addition, there are certainly cases outside of our current "number in the news" scope where one would not want to round all the way to one leading digit (e.g., in a detailed budgeting or accounting scenario, or in a scientific article being scrutinized by experts in the peer-review process). It is also likely that the cognitive benefits of rounding vary with the degree of rounding; for example, rounding a ten-digit number to five leading digits may not yield a similar benefit as rounding a six-digit number to one leading digit, due to diminishing returns. Future research might investigate a wider range of scenarios in which rounding might (or might not) be appropriate, as well as the marginal benefits of rounding to greater or lesser extents.

It would also be interesting to contrast the results of a such a study with the preferences for rounding that we uncovered in the two surveys conducted. For instance, is the degree of rounding that people prefer also the amount that would be optimal for the particular cognitive task at hand? If not, do people tend to prefer more or less rounding than the ideal amount? For instance, readers might say they want to see two or three leading digits, even though one leading digit might be optimal for approximate recall.

With regard to our survey findings, we uncovered relatively clear preferences for degrees of rounding; however, as mentioned above, our studies were not designed to reveal *why* people found various degrees of rounding acceptable or not. For instance, we currently cannot disentangle preferences based on absolute or relative accuracy (e.g., numbers should be shown within 5% of their precise values) from those based on aesthetics (e.g., having too few or too many digits simply doesn't look good to people). Additionally, in our surveys participants were first shown the original precise numbers and then asked to rate different rounding options. However, we might obtain different results if the original numbers were not shown – in such a scenario, participants might be unaware of the measurement precision and therefore less inclined to rate the highly rounded options as acceptable [17]. Therefore, future work could employ think-aloud tactics or other cognitive task analyses [25] to better uncover details of the mechanisms at play.

Towards promoting numeracy and enhancing numerical cognition, our research has demonstrated that using round numbers is an effective strategy when people engage in shallow processing of such numbers (i.e., read news articles that contain them). At the same time, prior work has reported that promoting deeper number processing, for example by employing perspective sentences [41] or decreasing presentation quality [44], can yield similar benefits. A natural next step is to examine how these two lines of strategy work in tandem, i.e., whether processing numbers deeply would complement or reduce the cognitive advantages of round numbers. More generally, identifying when and how each strategy can be most effective is a crucial component in the design of future tools for number comprehension assistance.

Finally, we consolidated our findings into an intelligent service for rounding numbers. Similar to the perspective engine by [41], this service can be embedded in a text editor or function as a browser plug-in, where its role is to suggest appropriate rounding strategies to enhance the user's number comprehension. The service is able to automatically detect numbers suitable for rounding and to dynamically select the best rounding strategy based on user preferences and the number magnitudes. The source code and demonstration of our service can be found at https://github.com/jhofman/round-numbers-chi2022.

ACKNOWLEDGMENTS

We thank Chip Heath for suggesting this research topic and for the helpful conversations.

REFERENCES

[1] 1982. Letters to the Editor. *The American Statistician* 36, 1 (1982), 64–67. https://doi.org/10.1080/00031305.1982.10482782
[2] Muhammad Idrees Ahmad. 2016. The magical realism of body counts: How media credulity and flawed statistics sustain a controversial policy. *Journalism* 17, 1 (2016), 18–34.
[3] Jennifer Allen, Baird Howland, Markus Mobius, David Rothschild, and Duncan J Watts. 2020. Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances* 6, 14 (2020), eaay3539.
[4] Randal C. Archibold. 2014. Hope Dwindles for Hondurans Living in Peril. *The New York Times* (2014). https://www.nytimes.com/2014/08/03/world/americas/hope-dwindles-for-hondurans-living-in-peril.html
[5] American Psychological Association et al. 2019. Publication Manual of the American Psychological Association, (2020). (2019).
[6] Pablo J Barrio, Daniel G Goldstein, and Jake M Hofman. 2016. Improving comprehension of numbers in the news. In *Proceedings of the 2016 chi conference on human factors in computing systems*. 2729–2739.
[7] Kamil Bartoń. 2020. *MuMIn: Multi-Model Inference*. https://CRAN.R-project.org/package=MuMIn R package version 1.43.17.
[8] Susana Bautista, Raquel Hervás, Pablo Gervás, Richard Power, and Sandra Williams. 2011. How to make numerical information accessible: Experimental identification of simplification strategies. In *IFIP Conference on Human-Computer Interaction*. Springer, 57–64.
[9] Arthur G Bedeian, Michael C Sturman, and David L Streiner. 2009. Decimal dust, significant digits, and the search for stars. *Organizational Research Methods* 12, 4 (2009), 687–694.
[10] Tom Blount, Laura Koesten, Yuchen Zhao, and Elena Simperl. 2020. Understanding the Use of Narrative Patterns by Novice Data Storytellers.. In *CHIRA*. 128–138.
[11] Coy Callison, Rhonda Gibson, and Dolf Zillmann. 2009. How to report quantitative information in news stories. *Newspaper Research Journal* 30, 2 (2009), 43–55.
[12] Arun Chaganty and Percy Liang. 2016. How Much is 131 Million Dollars? Putting Numbers in Perspective with Compositional Descriptions. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 578–587. https://doi.org/10.18653/v1/P16-1055
[13] Jacob Cohen. 1992. Things I have learned (so far).. In *Annual Convention of the American Psychological Association, 98th, Aug, 1990, Boston, MA, US; Presented at the aforementioned conference*. American Psychological Association.

[14] Edward T Cokely, Mirta Galesic, Eric Schulz, Saima Ghazal, and Rocio Garcia-Retamero. 2012. Measuring risk literacy: the berlin numeracy test. *Judgment and Decision making* (2012).

[15] Jill Cowan. 2019. Who Got Into the University of California? *The New York Times* (2019). https://www.nytimes.com/2019/07/23/us/uc-admissions-2019.html

[16] Stanislas Dehaene and Jacques Mehler. 1992. Cross-linguistic regularities in the frequency of number words. *Cognition* 43, 1 (1992), 1–29.

[17] Claudiu Dimofte and Chris Janiszewski. 2013. The Illusion of Lie Effect: the Suspicious Fluency of Round Numbers. *ACR North American Advances* (2013).

[18] A. S. C. Ehrenberg. 1981. The Problem of Numeracy. *The American Statistician* 35, 2 (1981), 67–71. http://www.jstor.org/stable/2683143

[19] Raymond W Gibbs Jr and Gregory A Bryant. 2008. Striving for optimal relevance when answering questions. *Cognition* 106, 1 (2008), 345–369.

[20] Gerd Gigerenzer and Ulrich Hoffrage. 1995. How to improve Bayesian reasoning without instruction: frequency formats. *Psychological review* 102, 4 (1995), 684.

[21] Irving John Good. 1968. Statistical fallacies. *International encyclopedia of the social sciences* 5 (1968), 292–301.

[22] Steven Harrison. 2016. Journalists, numeracy and cultural capital. *Numeracy: Advancing Education in Quantitative Literacy* 9, 2 (2016).

[23] Fabrice Hervé and Armin Schwienbacher. 2018. Round-number bias in investment: Evidence from equity crowdfunding. *Finance* 39, 1 (2018), 71–105.

[24] Jessica Hullman, Yea-Seul Kim, Francis Nguyen, Lauren Speers, and Maneesh Agrawala. 2018. Improving comprehension of measurements using concrete re-expression strategies. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[25] Robert JB Hutton and Laura G Militello. 2017. Applied cognitive task analysis (ACTA): A practitioner's window into skilled decision making. In *Engineering psychology and cognitive ergonomics*. Routledge, 17–23.

[26] Carel JM Jansen and Mathijs MW Pollmann. 2001. On round numbers: Pragmatic aspects of numerical expressions. *Journal of quantitative linguistics* 8, 3 (2001), 187–201.

[27] Ji Youn Jeong and John L Crompton. 2017. The use of odd-ending numbers in the pricing of five tourism services in three different cultures. *Tourism Management* 62 (2017), 135–146.

[28] Yea-Seul Kim, Jessica Hullman, and Maneesh Agrawala. 2016. Generating personalized spatial analogies for distances and areas. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 38–48.

[29] Yea-Seul Kim, Paula Kayongo, Madeleine Grunde-McLaughlin, and Jessica Hullman. 2020. Bayesian-assisted inference from visualized data. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 989–999.

[30] Michael Kimmelman. 2014. The Case Against a Mammoth Frick Collection Addition. *The New York Times* (2014). https://www.nytimes.com/2014/07/31/arts/design/the-case-against-a-mammoth-frick-collection-addition.html

[31] Manfred Krifka. 2007. *Approximate interpretation of number words*. Humboldt-Universität zu Berlin, Philosophische Fakultät II.

[32] Charles Livingston and Paul S Voakes. 2005. *Working with numbers and statistics: A handbook for journalists*. Routledge.

[33] J David Mason, Alice F Healy, and William R Marmie. 1996. The effects of rounding on memory for numbers in addition problems. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 50, 3 (1996), 320.

[34] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.

[35] MW McCall and Philip Bobko. 1990. Research methods in the service of discovery. *Handbook of industrial and organizational psychology* 1 (1990), 381–418.

[36] Academy of Management. 2007. Style guide for authors. *Academy of Management Journal* (2007), 472–475.

[37] European Association of Science Editors. 2011. EASE guidelines for authors and translators of scientific articles to be published in English. *J Teh Univ Heart Ctr* 6 (2011), 206–210.

[38] Ellen Peters. 2012. Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science* 21, 1 (2012), 31–35.

[39] Ellen Peters, Daniel Västfjäll, Paul Slovic, CK Mertz, Ketti Mazzocco, and Stephan Dickert. 2006. Numeracy and decision making. *Psychological science* 17, 5 (2006), 407–413.

[40] The Associated Press. 2020. *The Associated Press Stylebook*. The Associated Press.

[41] Christopher Riederer, Jake M Hofman, and Daniel G Goldstein. 2018. To put that in perspective: Generating analogies that make numbers easier to understand. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–10.

[42] Franklin E Satterthwaite. 1946. An approximate distribution of estimates of variance components. *Biometrics bulletin* 2, 6 (1946), 110–114.

[43] Herbert A Simon. 1996. *The sciences of the artificial*. MIT press.

[44] Frank Soboczenski, Paul Cairns, and Anna L Cox. 2013. Increasing accuracy by decreasing presentation quality in transcription tasks. In *IFIP Conference on Human-Computer Interaction*. Springer, 380–394.

[45] Stephanie Solt, Chris Cummins, and Marijan Palmović. 2017. The preference for approximation. *International Review of Pragmatics* 9, 2 (2017), 248–268.

[46] Dan Sperber and Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Vol. 142. Citeseer.

[47] Hal Stern and Thomas M Cover. 1989. Maximum entropy and the lottery. *J. Amer. Statist. Assoc.* 84, 408 (1989), 980–985.

[48] Manoj Thomas, Daniel H Simon, and Vrinda Kadiyali. 2010. The price precision effect: Evidence from laboratory and market data. *Marketing Science* 29, 1 (2010), 175–190.

[49] Jean-Baptiste Van Der Henst and Dan Sperber. 2004. Testing the cognitive and communicative principles of relevance. In *Experimental pragmatics*. Springer, 141–171.

[50] Hadley Wickham. 2020. *modelr: Modelling Functions that Work with the Pipe*. https://CRAN.R-project.org/package=modelr R package version 0.1.8.