Journal of Experimental Social Psychology 90 (2020) 103999

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Case Report

Truth from familiar turns of phrase: Word and number collocations in the corpus of language influence acceptance of novel claims

Y. Charles Zhang[a,*], Norbert Schwarz[b,c,d]

[a] School of Business, University of California, Riverside, United States of America
[b] Department of Psychology, University of Southern California, United States of America
[c] Dornsife Mind & Society Center, University of Southern California, United States of America
[d] Marshall School of Business, University of Southern California, United States of America

ABSTRACT

People are more likely to accept a claim as true, the more often they heard it in the past. We test whether using frequently encountered formal characteristics in constructing a novel claim increases its acceptance as true. A corpus analysis (study 1) established that, in everyday language use, lower-bound modifiers (e.g., “more than”) collocate more frequently with large numbers than upper-bound modifiers (e.g., “less than”). This regularity influences which numbers people expect to follow a given modifier (study 2): large (small) numbers are categorized faster and more accurately when primed by a lower- (upper-) bound modifier than an upper- (lower-) bound modifier. Novel quantitative claims that conform with these collocation patterns are more likely to be judged true (study 3), indicating that the collocation frequency of generic elements of quantitative expressions can influence the perceived truth of novel specific claims. Collocation frequency influences truth judgment even when participants know that the choice of number was arbitrary and based on their zip-code (study 4), suggesting that the effect does not depend on speakers' assumed communicative intent and the perceived informational value of the statements. We conclude that familiar turns of phrase can increase the acceptance of novel claims.

1. Introduction

As demagogues have known for millennia, repetition can turn a dubious claim into an accepted truth. Experimental research confirmed this insight: the more often people hear or read a claim, the more likely they are to accept it as true, independent of its actual veracity (for the initial demonstration, see Hasher, Goldstein, & Toppino, 1977; for a meta-analysis, Dechêne, Stahl, Hansen, & Wänke, 2010). We test whether demagogues can spare themselves some effort by constructing their claims with turns of phrase to which people have been frequently exposed in the course of daily life: Is the same substantive statement more likely to be accepted as true when its surface characteristics are familiar due to their frequency in the corpus?

To date, the influence of repetition on judgments of truth has been tested through repeated exposure to identical statements (Dechêne et al., 2010). However, several findings suggest that exposure to the specifics of the claim may not be necessary. Merely exposing participants to a topic (e.g., “A hen's body temperature”) can increase endorsement of a later specific assertion (e.g., “The temperature of a hen's body is about 104F”; Begg, Armour, & Kerr, 1985). Moreover, exposure to a specific statement (“Crocodiles sleep with their eyes open.”) can increase later acceptance of a substantively opposite statement (“Crocodiles sleep with their eyes closed”; Garcia-Marques, Silva, Reber, & Unkelbach, 2015), just as repeated warnings that a claim is false can increase its acceptance as true after a delay (Skurnik, Yoon, Park, & Schwarz, 2005). This presumably reflects that the preceding exposure facilitates fluent processing of the target claim, which, in turn, increases acceptance of the claim as true. Supporting a fluency account, any other manipulation that makes a claim easier to process – from the color contrast of the print font (Reber & Schwarz, 1999) to rhyme (McGlone & Tofighbakhsh, 2000) and accent (Lev-Ari & Keysar, 2010) – also increases perceived truth (for a review, see Schwarz, 2018).

In everyday life, a particularly powerful repetition manipulation may be the collocation of words in the corpus of natural language. Some words are more likely to collocate than others, resulting in different exposure frequencies for different word combinations. Theoretically, substantively equivalent statements should be more likely to be accepted as true when their wording includes frequently encountered word combinations than when it does not. We test this possibility by drawing on a class of quantitative expressions, known as one-sided intervals (Hohle & Teigen, 2018; Teigen, 2008), such as “more than X" or “less than X". A corpus analysis of contemporary English (study 1) shows that lower-bound modifiers – such as *more than* or *over* – are frequently used to modify large numbers in everyday language use, whereas upper-bound modifiers – such as *less than* and *under* – are frequently used to modify small numbers. Accordingly, statements of

* Corresponding author at: 900 University Ave., Riverside, CA 92782, United States of America.
E-mail addresses: charles.zhang@ucr.edu (Y.C. Zhang), norbert.schwarz@usc.edu (N. Schwarz).

https://doi.org/10.1016/j.jesp.2020.103999
Received 5 June 2019; Received in revised form 17 April 2020; Accepted 21 April 2020
0022-1031/

the form lower-bound modifier + large number or upper-bound modifier + small number should be processed more fluently than statements of the less common form lower-bound modifier + small number or upper-bound modifier + large number. Study 2 shows that this is the case: large numbers are processed faster and more accurately when preceded by a lower-bound modifier, whereas small numbers are processed faster and more accurately when preceded by an upper-bound modifier.

Of interest is whether claims that contain common combinations of modifier + number are more likely to be judged true than claims with less common combinations. Study 3 shows that this is the case. As detailed below, however, this observation may reflect that those more common modifier + number combinations (i.e., lower-bound modifier + large number or upper-bound modifier + small number) also result in claims that are more precise than the claims implied by the less common ones. If people assume that more precise estimates are more reliable (Zhang & Schwarz, 2012), they may overgeneralize this assumption and accept the more precise claims as true. We address this possibility in study 4 by informing participants that the choice of number was arbitrary and based on their zip-code. This manipulation undermines the assumption that the choice of the message is an intentional conversational act by a cooperative communicator, whose utterance is tailored to be informative for the purpose at hand (Grice, 1975). Nevertheless, participants were more likely to accept statements with common modifier + number combinations as true, replicating the results of study 3. In combination, these findings suggest that the frequency with which statement structures are encountered in daily language use can affect the perceived truth of claims that do or do not follow this structure, independent of the claim's specific content. From this perspective, creating truth through repetition does not require repetition of the same statement – using familiar turns of phrase is sufficient to increase the perceived veracity of a claim.

2. Study 1: corpus analysis

2.1. Method

We used the NOW (News On the Web) corpus (Davies, 2013) to examine the frequency of the collocations between lower- or upper-bound modifiers and numbers in the English language. In May 2018, when this analysis was conducted, NOW included about 6.0 billion words that appeared in web-based newspapers and magazines since 2010 in 20 countries.

Previous investigations of "one-sided intervals" focused on three pairs of modifiers (Cummins, 2015; Teigen, 2008): *more than-less than*, *at least-at most*, and *over-under*. In the NOW corpus, these six modifiers appear 2.9 million times and Appendix A.1 shows their respective frequencies. We counted the frequency of all unique numbers[1] that appear immediately after these modifiers. Appendix A.2 shows the top 10 occurrences for each modifier.

2.2. Results and discussion

The analysis identified $N = 26{,}930$ unique numbers that collocated with one of the lower-bound modifiers and $N = 4622$ unique numbers that collocated with one of the upper-bound modifiers. As expected, lower-bound modifiers were associated with larger numbers (*median* = 2304) than upper-bound modifiers (*median* = 128). A *t*-test on the log-transformed numbers confirmed that the difference is reliable ($M = 7.47$ vs. 5.21, $t(31550) = 32.8$, $p < .001$, $d = 0.37$); separate analyses confirmed this for each pair of modifiers (Table 1).

Table 1
Log-transformed means of the unique numbers following each modifier. Standard deviations in parentheses.

Lower-bound modifiers	Upper-bound modifiers	*p*-values for *t*-tests
More than	Less than	
7.7 (4.4)	*5.2 (4.4)*	*< 0.001*
Over	Under	
7.5 (4.5)	*5.3 (4.2)*	*< 0.001*
At least	At most	
7.2 (3.8)	*4.5 (7.0)*	*< 0.001*

Because some numbers occur more frequently in the corpus than others, we also conducted a weighted ANOVA that takes different frequencies of occurrence into account. Specifically, we compared the two groups of unique (log-transformed) numbers used in the preceding analysis weighted by the frequency with which each number occurs in the corpus. The results were comparable to the *t*-test ($M = 5.19$ vs. 3.09 for lower- and upper-bound modifiers, respectively; $F(1, 31{,}550) = 1219$, $p < .001$, $\eta^2 = 0.037$). The medians of the (untransformed) numbers that collocate with the lower- and upper-bound modifiers, weighted by their frequency of occurrence, are 90 and 18, respectively. Put simply, when we encounter an upper-bound modifier in everyday English, the associated number is less than 18 for half of the time; when we encounter a lower-bound modifier, the associated number is larger than 90 for half of the time.

3. Study 2: reaction time and accuracy

Theoretically, more frequently encountered combinations of modifier + number should be processed more fluently than less frequently encountered combinations. Hence, a large (vs. small) number should be categorized faster, and with higher accuracy, when primed by a lower-bound (vs. higher-bound) modifier. We tested this prediction in study 2. To ensure that a number without context can be inherently considered as large or small, we presented the numbers as percentages.

3.1. Method

This study was preregistered at http://aspredicted.org/blind.php?x=ip5t3c. Participants were 260 undergraduate students from a west-coast US university, which was the full number of students in a course-based subject pool. They were told that they would see a series of percentages and needed to judge whether each percentage was large or small. The instructions gave the example that "8% would be a small percentage number and 95% would be a large percentage number." Participants were then instructed to put their index fingers on the P and Q keys of the keyboard and to respond as fast as possible. Half of the participants were instructed to press Q if the number was small, and P if the number was large; the other half were assigned the opposite mapping.

The experiment was administered on PsychoPy 2.0 (Peirce, 2007). In addition to the four focal modifiers of interest *more than*, *less than*, *over*, and *under*,[2] we included four other modifiers – *about*, *exactly*, *precisely*, *roughly* – as controls. The control modifiers indicate neither an upper- nor a lower-bound. For each trial, one of the modifiers was presented for 200 milliseconds in the center of the screen. Immediately

[1] We only included numbers in their numeral form and ignored the units that follow the numbers; for example, "10 thousand" and "10 million" are both coded as 10. This was necessary because converting all units is neither feasible nor theoretically straightforward – do people think of 1 kg as 1000 g? However, a follow-up analysis of the unit "million" shows that it follows low-bound modifiers disproportionally more frequently than upper-bound modifiers. This suggests that an analysis that takes unit into account may identify even larger differences than currently reported in Study 1.

[2] In the remainder of the paper, we do not include the pair *at least/at most* due to the disproportionally low frequency of occurrence of the phrase *at most.*

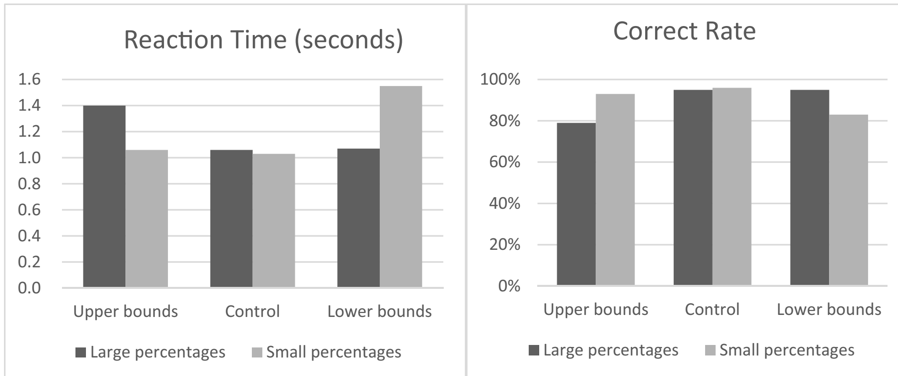

Fig. 1. Classification performance and reaction times by condition.

after that, a percentage number randomly chosen from 1%–15% or 85%–99% was presented in the same location on the screen using the same font. After participants pressed P or Q, the trial was over, and the next trial started immediately with a different modifier. Participants completed two trials for each modifier, 16 trials in total. The order of the presentation of modifiers was random for each participant.

For analysis, we categorized numbers between 1%–15% as small and numbers between 85%–99% as large. Of interest is whether participants classified a large number a) faster and b) more correctly after being exposed to a lower-bound than after being exposed to an upper-bound modifier, and vice versa for small numbers. For the present design, power analyses through PANGEA (Westfall, 2016) indicate a minimum detectable effect size of Cohen's $d = 0.17$ when contrasting upper- to lower-bound modifiers, a minimum detectable effect size of Cohen's $d = 0.14$ when contrasting upper- or lower-bound modifiers to control modifiers, and a minimum detectable effect size of Cohen's $d = 0.11$ when the analysis involves all modifiers, with power of 0.80. We report all measures, manipulations, and exclusions.

3.2. Results and discussion

3.2.1. Data preparation

Of the 260 participants, 31 misclassified the numbers more than half of the time. Their responses were entirely removed. Of the remaining $16 \times 229 = 3664$ responses, 26 responses had reaction times beyond three standard deviations from the mean. These specific responses were also removed. These exclusions follow the pre-registration protocol, leaving 3638 data points in the analysis. Of these, 911 were associated with the two lower-bound modifiers and 908 with the two upper-bound modifiers; 1819 were associated with the four control modifiers. The average reaction time was $M = 1.15$ ($SD = 0.96$) seconds, and the average correct rate was $M = 92\%$ ($SD = 10\%$).

For analysis, we constructed dummy variables for *bound* (1 = lower-bound modifiers; 0 = control modifiers; −1 = upper-bound modifiers) and *number* (1 = large focal number; −1 = small focal number). *Rtime* indicates reaction time (in seconds) and *Correct* indicates the correctness of the classification (1 = correct; 0 = incorrect).

3.2.2. Reaction time

We predicted that large numbers are categorized faster when they follow a lower-bound rather than upper-bound modifier; the opposite holds for small numbers. The results are consistent with this prediction (Fig. 1, left panel). To account for heterogeneity among participants and among individual modifiers within their categories (upper-bound and lower-bound), we ran a linear mixed model (Rtime = Bound + Number + Bound*Number + Participant + Modifier), treating *Participant* and *Modifier* as random effects.[3] This model allows for each participant to have generically different reaction times and for each modifier within categories to vary in fluency. Function *lmer* in R was used. The results (Table 2; for the full output, see Appendix B.1) revealed the predicted negative interaction effect ($b = -0.21$, $t(1690)$[4] $= -9.25, p < .001, d$[5] $= 0.20$): categorization was significantly faster when a large number was preceded by a briefly presented lower-bound modifier, or a small number was preceded by a briefly presented upper-bound modifier, than when the combinations of bound and number were reversed.

3.2.3. Accuracy

We further predicted higher accuracy when a large number follows a lower-bound rather than upper-bound modifier; the opposite holds for small numbers. The results are consistent with this prediction (Fig. 1, right panel). We ran a mixed logit model (with function *glmer* in R) to investigate correct rate (Correct = Bound + Number + Bound*-Number + Modifier + Participant). The predicted interaction is positive and significant ($b = 0.78$, $z = 8.50$, $p < .001$, OR[6] $= 2.19$): a larger number was more likely to be categorized correctly when it followed a briefly presented lower-bound modifier, and a small number was more likely to be categorized correctly when it followed a briefly presented upper-bound modifier, than when the combinations of bounds and numbers were reversed (Table 2; for the full output, see Appendix B.2).

An additional set of analyses took the control modifiers into consideration by contrasting participants' responses to numbers that followed the upper-bound or control modifiers, and to numbers that followed the lower-bound or control modifiers, respectively. The results (reported in Appendix C) were consistent with the above conclusions.

In sum, speed and accuracy were higher when large numbers followed lower-bound modifiers or small numbers followed upper-bound modifiers than when the modifier + number combinations were reversed. These findings are consistent with the assumption that more frequent exposure to a bound + number combination in the corpus of language (study 1) facilitates processing. Next, we test whether these fluency differences affect the likelihood that a substantive claim is accepted as true.

[3] Note that our conceptual rationale does not make predictions about between-level interactions (e.g., interactions of participants*bounds). Hence, we have not included random slopes. If one includes random slopes for exploratory purposes (following procedures suggested by Barr, Levy, Scheepers, & Tily, 2013), the models do not converge.

[4] The degrees of freedom were calculated based on Satterthwaite's approximation (Kuznetsova, Brockhoff, & Christensen, 2017)

[5] The effect size was calculated based on Westfall, Kenny, and Judd (2014).

[6] In the GLMMs throughout this paper, we report odds ratios (ORs) as indicators of the effect sizes of coefficients.

Table 2
Unstandardized coefficients of the models for the results of study 2. Standard errors in parentheses.***, *, +

	Reaction time	Correct rate
Bound	−0.041 (0.023)	−0.193* (0.088)
Number	0.044+ (0.022)	−0.037 (0.091)
Bound*Number	−0.206*** (0.022)	0.782*** (0.092)
Participants (variance)	0.291	1.384
Modifiers (variance)	0.004	0.000

*** $p < .001$.
* $p < .05$.
+ $p < .1$.

4. Study 3: truth judgment

Given the robust impact of fluency on judgments of truth (for reviews, see Dechêne et al., 2010; Schwarz, 2018), we predict that factual claims are more likely to be accepted as true when they entail one of the more fluently processed combinations of bounds and numbers (i.e., lower-bound + large number or upper-bound + small number) than when they entail one of the less fluently processed (i.e., reversed) combinations.

For example, "*More than* 80% of the mass in our solar system is composed of gases" should be more likely to be judged true than "*Less than* 80% of the mass in our solar system is composed of gases" and "*Less than* 5% of all tax returns are audited by the IRS" should seem more true than "*More than* 5% of all tax returns are audited by the IRS." Of course, different combinations of bound + number also differ in the substantive claims they make and may seem more or less true for that reason. Suppose, however, that the specific number used in the claim is the median estimate provided by the population from which the participants are drawn. In that case, claims of "*more than [median]"* and *"less than [median]"* should each be endorsed by half of the participants. Our rationale predicts that this is not the case. Whenever the median estimate provided by the population is a low number, a claim should be more likely to be accepted as true when it entails an upper bound ("*less than [median]")* than when it entails a lower bound (*"more than [median]")*. Conversely, whenever the median estimate provided by the population is a large number, a claim should be less likely to be accepted as true when it entails an upper bound ("*less than [median]")* than when it entails a lower bound (*"more than [median]")*. Study 3 relied on this strategy.

We used a pilot study to identify factual claims for which the median estimate is a high or low number. Using the median estimates from this exercise, we generated claims of the form upper-bound + median estimate or lower-bound + median estimate. We predict an interaction of bound and size of the median estimate in judgments of truth, such that (i) claims involving frequent (and hence more fluently processed) combinations (upper-bound + small median estimates and lower-bound + large median estimates) are accepted as true by more than half of the participants, whereas (ii) claims involving less frequent combinations (upper-bound + large median estimates and lower-bound + small median estimates) are accepted as true by less than half of the participants.

4.1. Pilot study and materials

Research assistants searched for factual statements in the form of "X % of A is B" where X% is either very large or very small. Examples include "0.86% of all tax returns are audited by the IRS" and "97% of all our planet's water is contained in the ocean," among others. We then asked 29 Amazon Mechanical Turk workers (each paid 20 cents) to estimate the numbers in those statements, for example, "__ % of all tax returns are audited by the IRS." The four statements with the largest median estimates and the four statements with the smallest median estimates (Appendix D) were selected for the main study.

4.2. Method

The traditional illusion-of-truth effect has an average effect size of Cohen's $d = 0.49$ (Dechêne et al., 2010). To find a difference in the truth judgment of a previously exposed versus unexposed claim, 134 participants are required to reach a power of 0.80 at a significance level of 0.05. Since the present research investigates truth judgment in situations where participants were never exposed to the exact claims, the effect size could be smaller. We therefore aimed to recruit a higher number of participants. This study was pre-registered at http://aspredicted.org/blind.php?x=wz45ck.

$N = 329$ Mechanical Turk workers (mean age = 38; 56% female) were presented all eight statements in random orders, each randomly paired with "more than," "less than," "over," or "under." For example, "More than 80% of the mass in our solar system is composed of gases." The percentage number in each statement was the median from the pilot study. Participants judged whether each statement was true or false. For the present design, power analysis through PANGEA (Westfall, 2016) indicates a minimum detectable effect size of Cohen's $d = 0.11$. We report all measures, manipulations, and exclusions.

4.3. Results and discussion

A total of 2630 (329 * 8 = 2632 minus two missing data points) "yes-no" responses were recorded. For analysis, we constructed a dummy variable *Number* (1 = questions with large percentage numbers [Q1 to Q4 in appendix D]; −1 = questions with small percentage numbers [Q5 to Q8]) and a dummy variable *Bound* (1 = questions starting with *More than* or *Over;* −1 = questions starting with *Less than* or *Under*). Participants' responses were labeled as *True* (1 = "true"; 0 = "false").

We fitted a mixed logit model to examine the difference between truth judgments pertaining to claims with frequent vs. less frequent bound + number combinations: True = Number + Bound + Number*Bound + Participant + Question. *Participant* and *Question* are random effects; all other variables are fixed effects. As predicted, participants were more likely to judge a statement as true when the statement conformed to frequent bound + number combinations than when it did not. This difference is reflected in a significant interaction of *Number*Bound* ($b = 0.50$, $z = 4.13$, $p < .001$, $OR = 1.65$). For the full output, see Appendix E. As shown in Table 3, statements of the form bound + median were accepted as true more than 50% of the time when they combined a lower bound with a large focal number ($P = 74\%$; $z = 5.17$, $p < .001$) or an upper bound with a small focal number ($P = 71\%$; $z = 4.13$, $p < .001$). In contrast, acceptance rates did not differ from 50% when the statements combined an upper bound with a large focal number ($P = 52\%$; $z = 0.964$, $p = .335$) or a lower bound with a small focal number ($P = 52\%$; $z = 0.392$, $p = .690$). Put simply, using frequent bound + number combinations increased acceptance as true beyond what would be expected based on the median, whereas using infrequent combinations did not decrease acceptance below what would be expected. Next, we turn to the possibility that conversational processes may contribute to the observed pattern.

5. Study 4: fluency or informational content?

In the case of large numbers, lower bounds (e.g., *more than 90%*) specify a narrower range of likely outcomes than upper bounds (e.g., *less than 90%*); conversely, upper bounds (e.g., *less than 5%*) specify a narrower range of likely outcomes than lower bounds (e.g., *more than 5%*) in the case of small numbers. In both cases, the more frequently used bound + number combination excludes more states of the world than the less frequently used combination and is hence more informative. This difference in informational value presumably drives the

Table 3
The average proportion of "True" answers to each question (study 3). Standard deviations in parentheses.

	Large focal number					Small focal number				
	solar system	train	water	gift cards	Average	crypto-currency	green eyes	immi-grants	IRS	Average
Lower bound	85%	67%	82%	62%	74% (10%)	42%	45%	59%	61%	52% (8%)
Upper bound	48%	48%	53%	60%	52% (5%)	79%	72%	60%	72%	71% (7%)

differential prevalence of these bound + number combinations in everyday life – the less informative combinations are less useful and hence less frequently used. Note, however, that claims with a narrower range are less likely to be true – in the absence of highly diagnostic information, it is more likely that the true value falls within the wide range specified by less prevalent bound + number combinations than within the narrow range specified by more prevalent ones. Despite this higher risk of being wrong, people are inclined to accept the narrower claims as true (study 3).

One reason why they may do so is that the combinations with the narrower range have been encountered more frequently (study 1) and are hence processed more fluently (study 2). This can increase their acceptance as true (study 3) as has been observed for other fluency manipulations (for a review see Schwarz, 2018).

Alternatively, the observed effect may reflect that people assume that the communicator follows the norms of cooperative conversational conduct (Grice, 1975; Levinson, 1983) and provides information that is truthful, relevant and clear. When people assume that the communicator is cooperative, they are more influenced by statements that follow a linguistic format that conveys higher precision (Zhang & Schwarz, 2012, 2013). For example, people infer that a project is more likely to be completed on time when the contractor estimates the required time in more fine-grained units (Zhang & Schwarz, 2012). This makes a promised completion time of 365 days seem more credible than one of 12 months or 1 year, despite their substantive equivalence. Such inferences from the linguistic form of the statement are eliminated when recipients doubt that the communicator follows the norms of cooperative communication, which calls the informational value of the precise utterance into question (Zhang & Schwarz, 2012, 2013). This is consistent with the general observation that normatively irrelevant information may appear as relevant when presented by an apparently cooperative communicator (for reviews, see Schwarz, 1994, 1996).

Study 4 provides a replication of study 3 and addresses the competing process hypotheses. Specifically, participants were told that the number selection is arbitrary and based on their own zip-code, thus undermining the informational value of the precision conveyed by the bound + number combination. This should eliminate or attenuate the effect observed in study 3 if it was due to the differential informational value of the claims. In contrast, this manipulation should exert little influence when truth judgments are based on the metacognitive experience of fluent processing.

5.1. Method

Participants were 261 undergraduate students from a west-coast US university (mean age = 22; 51% female), which was the full number of students in a course-based subject pool. They were randomly assigned to the conditions of a 2 (focal number: large vs. small) x 2 (number selection: arbitrary vs. potentially intentional) x 2 (bounds: upper vs. lower) mixed design where the first two factors were manipulated between-subjects and the last factor was manipulated within-subject. Participants were told that their tasks were to make true-false judgments on quantitative statements like "More than 81% of all birds are black."

The manipulation of the first two factors took advantage of the fact that most students live in communities where the five-digit zip-codes begin with 92. In the number-selection-arbitrary conditions, participants were told that the percentage number in each statement would be generated by using the zip-code of their current residence. Those assigned to the large-focal-number condition were asked to enter the first and second digits of their zip code, which ranged from 90 to 94 ($M = 91.9$, $SD = 0.6$); those assigned to the small-focal-number condition entered the second and third digits of their zip code, which ranged from 02 to 28 ($M = 23.2$, $SD = 5.0$). On the next page, participants were first reminded that the two zip-code digits they entered would be used in the following questions. Next, they were presented in random orders with either the four statements associated with large percentage numbers or the four statements associated with small percentage numbers used in study 3 (Appendix D), except that the numbers were replaced with the participant's own zip-code digits. Two of the statements used "more than" and the other two used "less than."

In the number-selection-intentional condition, participants responded to the same statements without first entering digits of their zip-code. Unbeknownst to these participants, the numbers they saw in the statements were also their own zip-code numbers, extracted from demographic information which all participants were asked to complete beforehand. These numbers ranged from 91 to 95 ($M = 92.0$, $SD = 0.8$) in the large- and from 00 to 51 ($M = 23.3$, $SD = 8.1$) in the small-focal-number condition. We report all measures, manipulations, and exclusions.

5.2. Results and discussion

We compared the actual zip-codes of participants in the number-selection-arbitrary conditions with the two digits that they entered and found mismatches for 12 participants, many of whom entered the first two digits when asked to enter the 2nd and 3rd digits. These participants were removed. We also removed one participant whose 2nd and 3rd digits of zip-code were 00. This left 248 participants for analysis.

Table 5 presents the aggregated averages across statements. We fitted the same mixed logit model as in study 3 with an additional fixed effect and its interactions – *Number selection* (1 = intentional; −1 = arbitrary). Replicating study 3, the interaction *Number*Bound* was significant ($b = 0.84$; $z = 6.12$, $p < .001$, $OR = 2.31$). Unrelated to the hypothesis, the coefficient of *Number* and the interaction *Bound*Number selection* were also significant (*Number*: $b = -0.35$, $z = -2.97$, $p = .003$, $OR = 0.70$; *Bound*Number selection*: $b = 0.33$, $z = 2.47$, $p = .014$, $OR = 1.40$). More important, the three-way interaction *Number*Bound*Number selection* was not significant ($b = -0.04$; $z = -0.28$, $p = .777$, $OR = 0.96$). For the full output, see Appendix F. Overall, participants were more likely to judge a statement as true when the statement conformed to frequent bound + number

Table 5
The average proportion of "True" answers (study 4). Standard deviations in parentheses.

		Large focal number	Small focal number
Number selection arbitrary	More than	59% (7%)	35% (21%)
	Less than	49% (5%)	65% (13%)
Number selection intentional	More than	67% (9%)	46% (17%)
	Less than	44% (6%)	60% (12%)

combinations than when it did not, regardless of whether the focal number was generated arbitrarily (based on the participant's own zip-code) or was selected by the speaker presumably with intention.

6. General discussion

Our results are consistent with the hypothesis that the frequency with which words and numbers collocate in the corpus of language influences processing fluency and the perceived truth of novel quantitative claims that conform with, or deviate from, frequent collocation patterns. In the corpus of English language, lower-bound modifiers are more frequently collocated with large than with small numbers, whereas the reverse holds for upper-bound modifiers (study 1). This regularity bears on Teigen's (2008) observation that recipients infer from a lower-bound (vs. upper-bound) modifier that the focal quantity is large (vs. small). These inferences are consistent with the collocation patterns in the corpus, allowing people to draw on shared implicit knowledge in the design and comprehension of bounded estimates.

This implicit knowledge influences which numbers people expect to follow an upper- or a lower-bound modifier: numbers are categorized faster and more accurately when the modifier + number combination follows the collocation conventions in the corpus (study 2). More important, quantitative claims are more likely to be accepted as true when their wording observes these conventions than when it does not (study 3). This effect does not depend on perceived communicative intent and is also obtained when the selection of the numbers is explicitly arbitrary and based on participants' zip-code (study 4). The latter observation renders it unlikely that the advantage of claims that follow the collocation conventions is based on their perceived higher informational value.

In combination, these results provide initial evidence that a claim's perceived truth can depend on the collocation frequency of generic elements in the corpus of language, independent of the substantive content of the claim. This observation differs from extant research into repetition-based truth effects in important ways. In contrast to the bulk of research on fluency and truth, the present studies do not entail a repetition of the target claim (Dechêne et al., 2010) nor the repetition of an almost identical sentence (Garcia-Marques et al., 2015) or a manipulation of the claim's perceptual features (Reber & Schwarz, 1999). Instead, fluency derives from previous exposure to quantitative claims of a generically similar structure, although people are unaware of the regularity and cannot articulate it. We surmise that prior to the present discussion most readers had not noticed that "more than" commonly goes with a large number, whereas "less than" goes with a small number. That the collocation of modifiers and numbers can bias truth judgments suggests that numerous other collocation phenomena can do so as well: whenever two elements frequently collocate in the corpus of language, claims that conform to their collocation pattern should have a truth advantage. This makes familiar turns of phrase an important tool of persuasion, confirming speech writers' intuitions (Neale & Ely, 2007).

Open practices

The experiments in this article earned open materials and open data badges as well as a pre-registration badge for transparent practices. The materials of studies 3 and 4 are located in Appendix D. The data are located at the ICPSR data repository: https://www.openicpsr.org/openicpsr/project/108163/version/V3/view/. The pre-registration of study 2 is located at http://aspredicted.org/blind.php?x=ip5t3c. The pre-registration of study 3 is located at http://aspredicted.org/blind.php?x=wz45ck.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2020.103999.

References

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.

Begg, I., Armour, V., & Kerr, T. (1985). On believing what we remember. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 17*(3), 199.

Cummins, C. (2015). *Constraints on numerical expressions. Vol. 5*. Oxford University Press.

Davies, M. (2013) Corpus of News on the Web (NOW): 3+ billion words from 20 countries, updated every day. Available online at https://corpus.byu.edu/now/.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review, 14*(2), 238–257.

Garcia-Marques, T., Silva, R. R., Reber, R., & Unkelbach, C. (2015). Hearing a statement now and believing the opposite later. *Journal of Experimental Social Psychology, 56*, 126–129.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Vol. Eds.), *Syntax & Semantics. Vol. 3*.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior, 16*(1), 107–112.

Hohle, S. M., & Teigen, K. H. (2018). More than 50% or less than 70% chance: Pragmatic implications of single-bound probability estimates. *Journal of Behavioral Decision Making, 31*(1), 138–150.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46*(6), 1093–1096.

Levinson, S. C. (1983). *Pragmatics.* Cambridge, UK: Cambridge University Press.

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science, 11*, 424–428.

Neale, T. A., & Ely, D. (2007). *Speechwriting in perspective: A brief guide to effective and persuasive communication.* Washington, DC: Congressional Research Service.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in python. *Journal of Neuroscience Methods, 162*(1–2), 8–13.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*(3), 338–342.

Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. *Advances in Experimental Social Psychology, 26*, 123–162.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods and the logic of conversation.* Hillsdale, NJ: Erlbaum.

Schwarz, N. (2018). Of fluency, beauty, and truth: Inferences from metacognitive experiences. In J. Proust, & M. Fortier (Eds.). *Metacognitive diversity. An interdisciplinary approach* (pp. 25–46). New York: Oxford University Press.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31*(4), 713–724.

Teigen, K. H. (2008). More than X is a lot: Pragmatic implicatures of one-sided uncertainty intervals. *Social Cognition, 26*(4), 379–400.

Westfall, J. (2016). *PANGEA: Power analysis for general anova designs.* (Unpublished manuscript).

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*(5), 2020.

Zhang, Y. C., & Schwarz, N. (2012). How and why 1 year differs from 365 days: A conversational logic analysis of inferences from the granularity of quantitative expressions. *Journal of Consumer Research, 39*(2), 248–259.

Zhang, Y. C., & Schwarz, N. (2013). The power of precise numbers: A conversational logic analysis. *Journal of Experimental Social Psychology, 49*(5), 944–946.