

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Number Comprehension Impacts Political Judgments

Brian Guay, Zach Davis, Megan DeLaunay, Arthur Charlesworth,

University of Richmond

David Landy

Indiana University and the University of Richmond

Keywords: political cognition, political judgment, cognitive psychology, numerical reasoning, quantitative reasoning

Abstract

Making political judgments often requires comparing large numbers of very different magnitudes, such as evaluating the impact of reducing a \$6.9 billion National Science Foundation budget by \$370 million. Recent research indicates that many American adults have an extremely poor understanding of the relative size of such large numbers, which play a critical role in discussions of public policy. A question that remains is whether people's political judgments can be modified when given accurate information about numerical magnitudes.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

The present study examines this question by assessing the effect of participants' understanding of the magnitude of large numbers on their evaluation of political scenarios involving large numbers and their attitudes toward government spending. In three experiments, half of participants were randomly assigned to receive a training treatment, during which the actual magnitude of 10 million in relation to 1 billion was demonstrated on a number line. The results indicate not only that understanding of large numbers can be improved by a short training session, but that political judgments involving these numbers can be affected by this entirely non-political treatment related to the comprehension of large numbers.

“It may appear to be costly, \$11 million out of a \$7 billion funding for the National Science Foundation, but I think that however expensive an education may be, ignorance will probably cost our country more”

- Congressman Chaka Fattah (Congressional Record 2012)

Number Comprehension Impacts Political Judgments

The magnitude words *million*, *billion*, and *trillion* are commonplace in current U.S. political dialogue and decision-making. Voters must make sense of values that cross these scales,

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

particularly when members of Congress propose budget and deficit reductions. The news media often reports information involving large numerical values, such as that there are about 314 million people in the United States, the federal deficit for 2013 was \$683 billion, the national debt is \$17 trillion, and the National Security Agency—whose extensive data collection has been the topic of lively political debate—plans to store over a trillion trillion (i.e., 1,000,000,000,000,000,000,000,000) bytes of information at a data center in Utah (Utah Data Center 2013).

The question of how voters assimilate (especially quantitative) political information has long been a topic of importance in political science (Fiske et al. 1983; Lodge and Hamill 1986; Rahn 1993). As Charles Taber observed, “the engines of public opinion are individual citizens as information processors” (Taber 2003). In the last two decades there has been a considerable amount of attention specifically on what the public understands about politics and public policy (Delli Carpini and Keeter 1996). More recent research suggests that individuals are not only informed or uninformed, but also often misinformed. One of the most dramatic instances of misinformation is the public’s understanding of government spending. When the American public was asked how much of the 2010 U.S. budget was dedicated to foreign aid, the median response was 25%, a great overestimation of the actual 1% spent on foreign aid (World Public Opinion 2010). While the accuracy of this estimate improved with participants’ increased level of education, the median college educated participant still estimated that 15% of the U.S. budget went to foreign aid. More recent studies have found that the public also overestimates non-budget figures, such as the size of immigrant and minority populations (Gilens 2001; Sides and Citrin 2007). Several experiments have examined the effect of providing correct political information, such as the actual proportion of immigrants living in the U.S., on participants’

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

political beliefs, such as Americans' attitudes towards immigrants (Gilens 2001; Kuklinski et al. 2000; Lawrence and Sides 2011; Nyhan and Reifler 2010; Sides and Citrin 2007). Gilens (2001), for example, found that giving participants correct information about the amount the government actually spends on foreign aid decreases their desire to reduce its budget. These studies have led to the belief that the incorrect use of numbers in politics (*political innumeracy*) is primarily caused by a lack of correct information rather than an inability to use available information (Sides and Citrin 2007).

While it might seem rational to make use of relevant magnitudes when they are made available, there are several non-exclusive reasons people might not. One possibility is that people simply ignore quantitative information when making decisions (Sides and Citrin 2007). Another is that rational local interests lead people to bend such numerical information to substantiate their prior political judgments (Kahan et al. 2013; Kuklinski et al. 2000). Finally, it might be that people lack the cognitive tools to process numerical representations, or make specific errors in that processes. In this paper, we explore this last possibility in the case of representations of *large numbers* (defined here as numbers between about 10 thousand and 10 trillion).

We will suggest that when people do use quantitative information to reason, innumeracy leads people to misuse that information in particular ways. The experiments presented here probe role of numeracy by measuring changes to evaluations of number-based political scenarios after training designed to selectively improve number knowledge. Rather than simply providing voters with correct information, we seek to change voters' perceptions of information by increasing the accuracy of the magnitudes that they associate with large numbers. Another main departure from previous work in political number comprehension is our use of hybrid number notation (e.g. 2 million), rather than percentages (Landy et al. 2013). Although percentages more effectively

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

convey relative magnitude and long-form notation (2,000,000) is used in federal budgets, the news media that voters are most often exposed to predominantly use hybrid notation for numbers in the millions, billions, and trillions.

In the studies reported here, we will explore how the psychology of numerical reasoning affects evaluations of political scenarios (Experiment 2) and relates to voters' support of government programs (Experiment 3). We will also evaluate how numerical training affects, in the short term, these judgments (Experiments 2 and 3). In order to make clearer the relationship between numerical reasoning and voter judgments, we next briefly review results of relevant psychological research on numerical reasoning.

Number Line Estimation and the Psychology of Large Number Comprehension

When we think about large numbers, a common intuition is that they are all 'just big'. In the 18th century, Daniel Bournoulli (1954) captured this intuition with the assertion that perceptions of relative monetary quantities are *logarithmically compressed*: the experienced or perceived magnitude of a value is proportional to the logarithm of its actual value. This compressive approach forms the foundation of expected utility theory, and has become common in psychophysics under the name the *Weber/Fechner law*, or logarithmic compression. Because the logarithm grows more slowly than numerical value, two big numbers with a particular arithmetic difference feel closer together than do two small numbers with the same arithmetic difference. Put the other way, it takes a larger arithmetic difference to make two larger numbers 'feel' different in size that it does two smaller numbers. According to the Weber/Fechner law, the difference between 1 and 100 feels about equal to the difference between 10 million and 1 billion, because the relative difference is 100 in each case—even though the absolute difference

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

is 99 versus 990 million. More recently, power laws (Stevens 1975; Spence 1990; Barth and Paladino 2011; Cohen and Blanc-Goldhammer 2011) and other formulae have often been used instead of logarithms, but the general concept is the same; small values are sometimes more differentiated psychologically than are larger values over some range (see also Stevens 1975).

The number line task is a standard measure used by cognitive psychologists to assess individuals' understanding of magnitudes (Siegler and Opfer 2003, Opfer et al. 2011; Link et al. 2011; Cohen and Blanc-Goldhammer 2011; Dehaene and Brannon 2011; Feigenson et al. 2004; Siegler and Booth 2004; Ramani and Siegler 2008). This task has already been shown to relate to a number of mathematics outcomes in children, including predicting fraction learning (Bailey et al. 2012) and future mathematics achievement (Fazio et al. 2014). In this task, participants are asked to coordinate between a number line and another numerical representation—in our case, a number expressed in the hybrid format (e.g. 3 million). In the standard number line production task used here, participants are presented with a straight horizontal line, bounded by two numerical endpoints, and are asked to produce an estimate of the position on the line that corresponds to the given number (see Figure 1). An accurate, or *linear*, estimate is one where the position of the mark is proportional to the proportion of the placed number in the given range; when placing 3 on a line from 0 to 10, a linear estimate would be the 30% of the way from the left edge, because 3 is 30% of 10. A person with *compressive* responses might place the same number over 50% of the way from the left edge, suggesting that they overestimate small numbers relative to large ones.

Behavior on the number line task has been examined across several numerical ranges and across age groups ranging from pre-schoolers to adults (Barth and Paladino 2011; Barth et al. 2011; Booth and Siegler 2006; Siegler and Booth 2004; Siegler and Opfer 2003). The majority of

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

research involving the number line task examines children working with numbers under one million. Although an appropriate theoretical explanation of the error patterns found in participants' number line task performance is actively debated (Barth and Paladino 2011; Opfer et al. 2012; Rips et al. 2008), there is a consensus that children shift over developmental time from heavily compressive number line placements to relatively linear placements. This shift from overestimating small magnitudes to expressing them more accurately depends on the number range as well as a child's age (and presumably other demographic factors; see Barth et al. 2011; Barth and Paladino 2011; Opfer et al. 2011; Siegler and Robinson 1982). Young children start out producing compressive number line placements for all number ranges testable; slightly older children produce linear lines only for numbers up to about 10. Later (around 2nd grade), children produce linear estimates for numbers up to about 100. Tested populations of adults produce linear placements up to roughly 100 thousand (Thompson and Opfer 2010).

One might conclude from prior research that adults generally treat number lines linearly; however, this seems to be the case only when the numbers stay relatively small. Landy et al. (2013) extended the number line task to include numbers greater than 100 thousand, including those in the millions, and instructed participants to place numbers on a number line with a right hand endpoint of 1 billion. Similarly, Landy et al.'s participants correctly placed numbers within each subrange: 500 thousand lay halfway between 1 thousand and 999 thousand. Similarly, numbers in the millions were placed linearly between 1 million and 1 billion; 300 million was halfway between 200 million and 400 million.

Although participants could accurately place numbers in the separate thousands and millions ranges, the boundary between these two scales was systematically incorrectly placed, often by a large amount. About 40% of participants overestimated the point that they associated

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

with 1 million, placing it just before the halfway point on the line rather than just beyond the left end point (see Figure 2). Because it takes one thousand millions to make one billion, 1 million should properly be located on the extreme left-hand end of the line. This led to a systematic and distinctive error pattern, which was like a line broken into two connected pieces, or a ‘hockey stick’ (see Figures 2 and 3). Perhaps surprisingly, the breakpoint or ‘heel’ of the hockey stick was always associated with the value *1 million*. To the left of the point that a participant associated with 1 million, numbers in the thousands range were placed linearly relative to one another. Numbers in the millions range were also placed linearly relative to each other, to the right of the 1 million mark. Thus, on each side of 1 million participants placed numbers proportionally, in the correct linear pattern. However, because participants greatly overestimated the proportional location of 1 million, and because this value grounded their linear estimates elsewhere, their estimates across the entire range were heavily overestimated. As discussed below, this distinctive pattern means that responses across the entire range can be predicted very precisely using just a single parameter—the estimated location of one million on the line. Because this was almost always an overestimation, we call it the *overestimation value* (or sometimes just *overestimation*). We use this overestimation value as a summary for participants’ tendency toward compression on the number line task.

Although prior research indicates that a large proportion of the U.S. population makes systematic errors in number line placements, the relationship between number lines and political behavior is currently not well understood. The present studies contribute to this understanding in two ways: First, we explore the role of training in ameliorating errors in number comprehension. Although this kind of number line training has been successfully employed with small numbers (Thompson and Opfer 2010; Siegler and Opfer 2003), the errors with large numbers that we seek

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

to correct are unique because that they are made by adults, show hockey stick rather than logarithmic patterns (Landy et al. 2014; Resnick and Shipley 2012), and seem to originate in culturally specific methods of writing and talking about large magnitudes. Second, Experiments 2 and 3 explore the relationship between large number understanding and the evaluation of real political scenarios and statements of current relevance to the population studied. Landy, et al. (2013) also asked participants to evaluate political scenarios, but these scenarios were far from those that voters actual encounter. In that study, participants evaluated the actions of a fictional government involving large numbers. Participants whose number line placements indicated overestimates of numbers in the thousands and millions evaluated political scenarios involving large numbers differently than did participants with more linear placements, regardless of political party affiliation. Unfortunately, the political scenarios they used involved fictional currency and countries, and the stories often portrayed political scenarios that vary greatly from what U.S. voters are faced with, such as a national government deliberately taking steps to increase immigration.

The current study explores the effect of an entirely non-political treatment on participants' evaluation of actual, rather than fictitious, political scenarios described by modified news media excerpts. News media has been shown to have strong effects on voter behavior and is a large source of the public's political knowledge (Druckman 2004; Iyengar and Kinder 1987). By evaluating both participants' estimation of large numbers and evaluations of political scenarios before and after this treatment, we are able to determine an effect of improved number processing on political evaluations. While previous studies have attempted to change political attitudes by providing correct factual political information (Gilens 2001; Lawrence and Sides

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

2011; Nyhan and Reifler 2010; Sides and Citrin 2007), we seek to modify how people may process political information.

In Experiment 3 we examined how large-number numeracy relates to support for increased government spending on foreign aid and National Science Foundation research, which made up 1.5% (\$52 billion) and .15% (\$5 billion) of the federal budget in 2013, respectively. Since people who interpret values compressively are predicted to systematically overestimate the proportions—even when given accurate numerical values—we expect people with more linear estimates to be relatively more supportive of increasing funding to small programs such as these. We also expose people to training, to see whether short-term shifts in number line behavior shift support for government spending on these programs.

Common Methods Across Experiments

Mechanical Turk

In all three experiments, we use participants recruited from Amazon’s Mechanical Turk (MTurk), which has become increasingly accepted as a data collection method in psychology and cognitive science, but has been slower to gain popularity in political science. MTurk is an online “marketplace” in which Internet users can select and participate in Human Intelligence Tasks, such as behavioral science studies, for small monetary compensation (Mason and Suri 2012). MTurk allows researchers to access larger populations at a lower cost than more traditional data collection methods. A number of studies that have compared data collected on MTurk with NES online surveys, NES in-person surveys, and in-person convenient samples experiments have shown that MTurk samples are often just as representative of the general U.S. population as other methods (Berinsky et al. 2012; Paolacci et al. 2010). While many studies in the social

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

sciences involving experimental design use undergraduate students, this population is known to be less diverse and to have developmental and socio-political differences from the general population (Berinsky et al. 2012; Buhrmester et al. 2011; Henry 2008; Visser et al. 2000).

Particularly concerning for political scientists, student populations tend to be younger and more liberal than the national population (Henry 2008). In the field of political behavior, MTurk has recently been used as the sole population in experiments (Krupnikov and Bauer 2013; Crawford and Pilansky 2012; Travers 2013) or to run supplementary analyses (Gerber et al. 2012; Fausey and Mattlock 2012; Schaffner 2011).

In Experiments 2 and 3, we used an instructional manipulation check (Oppenheimer et al. 2009) to identify and eliminate participants who submitted responses without reading the provided stimuli. The content of these manipulation checks is described below. We also ensured that each MTurk account only participated in each study once and politely refused participants who had already participated. We were also able to ensure that participants could not participate in more than one of our experiments¹.

¹ To ensure that the same MTurk account could not complete our study twice, we used a python script that registered each MTurk participant's unique "workerID" number, which is associated with their MTurk account.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Experiment 1**Method****Participants**

A total of 300 participants were recruited from MTurk after self-selecting the study, which was run in November 2011. Participants saw a brief description of the study before participating and were compensated 25 cents for roughly 5 minutes of their time². 51% percent of participants identified as female, with an average participant age of 34. Participants were required to be of at least 18 years of age and live in the United States.

Procedure

Experiment 1 sought to determine whether a simple training treatment could improve numeracy. We followed Landy et al. (2013) in giving participants sixteen number line placements on a number line ranging from one thousand to one billion (see Figure 1; exact stimuli can be found in the online appendix). When people encounter large magnitudes presented by politicians or in the news media, most probably rely on implicit understandings of magnitudes rather than explicit calculations. The use of one thousand, rather than zero, as the left endpoint was intended to discourage participants from explicitly calculating percentages, though data from Experiment 3 and pilot studies suggest that there is little difference in number line placements when using zero as the left endpoint. The stimuli and endpoints were both presented in hybrid notation (e.g., as 1 billion rather than 1,000,000,000) to increase external validity. After the first

² The description that participants saw before selecting the study was: ² “Answer some questions about numbers. You will be asked to make a few simple decisions about some numbers, without using a calculator or other tool”.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

eight placements, half of the participants (control condition) read an encouragement text, thanking them for their effort and asking them to continue trying their best. The other half of participants (training condition) saw a screen stating that 1 billion is equal to one thousand millions and indicating the correct placement of 10 million on a line from 1 thousand to 1 billion. All participants then completed eight additional number line placements. After this, participants completed a set of fictional scenario evaluations similar to Landy et al. (2013). Because the focus of this paper is on plausible political scenarios, these results will not be further discussed here; it should be noted, however, that they are in no way contrary to any of the conclusions reached here.

Analysis

Based on prior research, we predicted that participants' number line placements would follow a 'broken line' or 'hockey stick' pattern, in which number line placements of values in the thousands range are spaced linearly to the left of the overestimation value, and values in the millions range are spaced linearly to the right of the overestimation value. By using this model, a participant's behavior can be accurately summarized by where the overestimation value is located on the line: equivalently, where did the participant "place" the value 1 million? 1 million was intentionally not included as one of our stimuli, because (1) the numerical simplicity of 1 million might induce special reasoning processes not common with large number comparisons generally, (2) because that value lies at a boundary and might induce learning, thus artificially influencing results (Landy et al. 2014), and (3) responses across the full range are of primary interest here. Instead, we use an estimate of the position of 1 million made by finding the best fitting overestimation value based on where linear estimations of thousands-range numbers and millions-range numbers diverge as a guide to performance across the range. For a perfectly

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

numerate participant, this overestimation value would be equal to 1 million, indicating that the participant accurately placed 1 million at the actual location of 1 million on the number line (In this case the participant is not actually overestimating, but we retain the term “overestimation value” to remind the reader that these values were in fact almost always overestimates). For a highly innumerate participant, this overestimation value might be 400 million, nearly halfway across the number line.,

Calculating the overestimation value is far more informative than simply calculating the mean of a participant’s number line placements (though the two are highly correlated), because the errors that innumerate participants make are not evenly distributed across the number range: both highly numerate and highly innumerate participants, for example, are likely to place “980 million” in the same place on the number line—far to the right. Simply calculating means for each number line placement does not capture the full weight of the difference between numerate and innumerate participants; modeling these participants in terms of the location of the 1 million ‘break’ in the hockey stick sensitively captures the pattern of errors produced.

Results & Discussion

For each participant, the eight number line placements were used to find the best-fitting subjective estimate of 1 million (the *overestimation value*) by fitting a segmented linear model. This model has one free parameter: the subjective value of 1 million (relative to 1 billion and 1 thousand). Behavior on either side was fitted as linear. This model is capable of matching both numerate linear behaviors (if the overestimation value is located very close to the normative location of 1 million) and hockey stick patterns (when it is not), as discussed above. Models were fitted to data using the *optim* function in R (Bates et al. 2013), minimizing the squared error of

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

the model fit to produce a maximum likelihood estimate of the subjective value of 1 million.

This estimate was made twice for each individual participant: once for number line placements made before the treatment, and once for number line placements made after, and the two values were fitted independently.

To evaluate the effect of the condition on training, we ran a linear regression on post-treatment overestimation values. The regression included a factor for treatment type (control or training) and included pre-treatment overestimation as a covariate. These values must be non-normal, because they are bounded between 0 and 1; they were clearly multi-modal by visual inspection: significance was evaluated using a permutation test with 10,000 replications. In all cases, the direction of the effect was theoretically predicted, so one-sided significance tests were used. Evaluation of pre-treatment overestimation values revealed no differences (Mean(control)=210 million, Mean(training)=200 million, $p = 0.59$). Pre-treatment overestimation values did significantly predict post-treatment overestimation ($p < 0.001$). The treatment had a statistically significant and moderately sized impact on overestimation (see Figure 4; $p = 0.001$; unstandardized $\beta = 60$ million). The estimated effect size (using Cohen's d) for the training treatment on overestimation was 0.37; that of the control condition was 0.004.

Experiment 1 demonstrates that a simple, direct training treatment instructing participants on the value of a single point on the number line serves to substantially improve the estimated subjective value of 1 million. Experiment 2 aims to evaluate whether learning about numbers on the relatively artificial number line task might impact a participant's evaluation of political scenarios.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Experiment 2**Method****Participants**

A total of 374 participants were recruited from MTurk after self-selecting the study, and were compensated 25 cents. The study was conducted in May, 2013. As in Experiment 1, only participants who were at least 18 years of age and resided in the United States were eligible to participate in the experiment. MTurk users were able to see a brief description of the study (“Answer some questions about numbers and politics) before participating. Forty-six percent of participants identified as female and 82% as white, with an average age of 33 years. Fifty-nine percent of participants identified as either a Strong Democrat, Weak Democrat, or Independent leaning Democrat, 14% as Independent, and 27% as a Strong or Weak Republican or Independent leaning Republican. Our sample is slightly younger and more liberal than the adult convenience samples reported by Berinsky et al. (2012).

Procedure

Experiment 2 replicates the structure of Experiment 1 by asking participants to complete a similar number line task before and after a treatment. As with Experiment 1, the training treatment condition received correct numerical information while the control treatment group was instructed to take a 30-second break. The unique component of Experiment 2 is the inclusion of the qualitative evaluation of political scenarios before and after the treatment, as well as the collection of political demographic information. Figure 5 depicts the organization of tasks in the study.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Demographic Questions

Participants first answered a set of basic demographic questions, including measures of educational attainment, household income, self-reported political party affiliation, attention to politics, and political participation³. They then answered the commonly used five general political knowledge questions recommended by Delli Carpini and Keeter (1996) to measure prior political knowledge⁴. Pilot studies indicated that participants had trouble differentiating the deficit and the debt, and so participants in this study were asked to differentiate the definitions of the deficit and debt, and were then shown the correct answer to establish an equal understanding prior to evaluating political scenarios involving these terms.

Political Scenario Evaluations

Participants then rated the impact of a series of political scenarios taken from real news sources. Half of participants saw political scenarios involving large numbers (i.e. statements about a \$17 billion spending cut), while the other half of participants were given scenarios that had no quantitative information (i.e. statements about President Kennedy's assassination).

Participants rated both types of scenarios on an 8-point scale ranging from “no impact” to “a huge impact”. Scenarios involving numerical information asked a participant to rate the impact of a spending cut on a greater budget or deficit, while the non-numerical scenarios asked participants to rate the impact of an event, such as the JFK assassination, on an arbitrary value

³ Political party affiliation was measured using the ANES seven-point scale.

⁴ The five questions recommended by Delli Keeter and Carpini are: 1) What job or office is currently held by Joe Biden? 2) Whose responsibility is it to determine if a law is constitutional? 3) How much of a majority is required for the U.S. Senate and House to override a presidential veto? 4) Which party had the most members in the House of Representatives in Washington before the 2012 election? 5) Which party is more conservative at a national level?

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

such as the stability of the nation. Our instructions to rate the ‘impact’ of a spending cut or event were intentionally vague. A participant’s impact rating of a spending cut, for example, should incorporate some combination of factors, including the importance the participant places on a balanced budget, the program or department being cut, and, by our hypothesis, the participant’s level of numeracy. Moreover, although participants always had a larger budget or deficit to compare a spending cut to, some cuts are objectively larger than others and a discerning participant might look beyond the local context and consider all cuts in terms of the entire federal budget.

By including non-numerical political scenarios, we were able to test whether the training treatment, which provided correct numerical information, had a selective effect on only political scenarios that included these large numbers. Although impact estimates might be affected by numerical training—say if number training shifts attention toward the lower end of scales—if people utilize quantitative information in evaluating political scenarios then numerical estimates should be selectively more affected.

Although we cannot speak to accuracy of these impact ratings, since there is obviously no objectively correct answer to the ‘impact’ of a budget cut, we can examine differences on impact ratings between numerate and innumerate participants, those who received the training and control treatments, and those who received the numerical and non-numerical scenarios.

Number Line Placement

Participants next completed a number line task that was identical to the one used in Experiment 1 to test their pre-treatment numeracy level. The same instructions were used, and the same set of numbers was used.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Treatment and Post-treatment Measures

Half of participants received the training treatment from Experiment 1 and half received the control treatment. After this, participants completed a post-treatment number line task and set of political scenario evaluations.

An instructional manipulation check was included after the demographic questions in order to determine which participants were carefully reading the full text of our study. In this trial, a political scenario was presented to participants as usual, but the final sentence asked participants to choose the third answer regardless of the question. This was a deliberately difficult manipulation check, and only 57% of participants passed it. The intention of this question was to ensure that selected participants were fully paying attention; participants who did not correctly answer the question were eliminated from the study.

A few specific predictions can be made. We expect to replicate the findings of Experiment 1 that indicate that a short training treatment can improve numeracy. With regard to the political scenario evaluations, we expect that participants who demonstrated higher levels of numeracy on the pre-treatment number line task will have lower impact ratings of pre-treatment political scenarios involving numerical information. We also expect that participants who receive the training treatment will have lower post-treatment impact ratings.

If people use the numerical information at all, then larger proportions of spending cuts to budgets/deficits should receive larger responses, all other things (such as political ideology) being equal. In this experiment, these proportions were always quite small (less than 5% of the total). If these evaluations are mediated by how participants understand relative numerical proportion, a participant with a highly accurate estimate of numerical values should rate an impact no higher than would an otherwise identical participant with a highly segmented number

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

line estimation pattern. Thus, teaching participants the accurate location of numbers on number lines should negatively impact their estimates of impact, regardless of other factors. This clear and general prediction forms the core hypothesis of this study.

Results & Discussion

Number Line Placement

Number line placements were analyzed as in Experiment 1. Number line placements were analyzed by a linear regression using pre-treatment subjective value of 1 million relative to 1 billion (the overestimation value) values as a covariate, and both condition factors (treatment type and scenario type) as independent variables. Recall that the overestimation value captures the primary error pattern in responses, and is our major measure of large-number numeracy. Statistical significance was estimated via a bootstrap with 10,000 random permutations. A preliminary analysis of pre-treatment overestimation values confirmed that there were no significant differences across groups prior to training (all p 's > 0.2).

As in Experiment 1, post-treatment overestimation values were significantly affected by the treatment type, such that participants who received the training treatment gave lower, more accurate responses ($p = 0.0001$; unstandardized $\beta=100$ million; see Figure 6). Since numerical task was identical across both judgment conditions, one does not expect it to affect overestimation values, and it didn't: there was no main effect of scenario type ($p=0.67$), and no interaction between scenario type and training ($p=0.24$) on overestimation.

Scenario evaluations

Raw scenario evaluation scores were unsurprisingly heavily skewed by the particular scenario provided, which contained different proportions and political content. Because the

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

relative value of the evaluation and not the raw evaluation itself was of interest here, responses to each question were separately converted to z-scores, and the resulting normalized scenario evaluations were analyzed.

Pre-treatment normalized scenario evaluations were submitted a linear regression using scenario and treatment conditions as predictors, to allow us to detect violations of random assignment that might impact subsequent analysis. There was no difference among the randomly assigned groups at pretest (all p 's > 0.5).

Post-treatment scenario evaluations were submitted to a separate linear regression over condition and judgments time, using pre-treatment scenario evaluations and million overestimation values as covariates, and again estimating significance using a bootstrap analysis. There was a significant and substantial effect of treatment type on evaluations, ($p = 0.0044$, $\beta = 0.42$) Training had a selective impact on numerical judgments (interaction $p = 0.046$, for word judgments Cohen's $d = 0.10$; for numerical judgments $d = 0.53$; see Figure 7). Training in number line placement selectively reduced scenario evaluations when those evaluations were numerical in character.

Raw scenario evaluation scores are, as described earlier, of limited value in this study. Many factors specific to the narratives and to the views of the individual participants are likely to affect impact ratings, and the decision to use real narratives made it undesirable to counterbalance specific numerical values across framings. Furthermore, in proportional terms, all cuts were quite small (less than 5%). Nevertheless, as a check that people were at least somewhat sensitive to the numerical values used in the political scenarios, we evaluated, at pre- and post-test, whether larger numerical values were at least associated with larger impact estimates. A bootstrapped linear regression indicated positive effects of the numerical value of the budget

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

item being cut at pretest ($p \sim 0.04$) and posttest ($p \sim 0.03$). The proportion of the budget item being cut and the larger budget were significantly positively related to judgments at posttest ($p < 0.001$), but were non-significantly negatively related at pre-test ($p \sim 0.45$). Though it would be tempting to take this as evidence that number line training and experience improved sensitivity to political numerical values, we caution against this judgment, because of the many differences between political scenarios mentioned above. However, the results are reassuring enough to proceed with the primary analyses of interest.

Individual differences analyses

Individual differences in pretest behavior were conducted to evaluate how particular groups might differ in initial scenario evaluation. Two linear regressions using age, gender, political party identification, political knowledge, education level, and income as independent variables predicting overestimation of 1 million on the number line in the first regression, and normalized scenario evaluation in the second (see Table 1). The entire sample was included for number line placement, but only the numeric judgment group was included for scenario evaluations. After Bonferroni correction for multiple comparisons, gender emerged as the only significant predictor of numeracy; male participants were on average more numerate than female participants. Scenario evaluations did not differ along any demographic or individual factor, except political knowledge: people with better political knowledge gave lower estimated ratings to the impact of budget cuts.

Although very few significant predictors of number line behavior emerged from this sample, the sample is quite small. Experiment 3 includes a much larger sample, which should make it more possible to evaluate any significant demographic differences.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Table 1: Regression analysis of individual demographic factors in predicting overestimation value and political scenario evaluation.

Variable	Overestimation Value		Scenario evaluations	
	β (millions)	95% CI	β	95%CI
Education	-10	[-0.02, -0.13]	0.0	[-0.10, 0.10]
Income	-8	[-0.02, 0.009]	-0.024	[-0.16, 0.11]
Age	-1	[-0.003, 0.001]	-0.009	[-0.03, -0.01]
Gender=Male	-100 ***	[-0.16, -0.05]	-0.39	[-0.80, -0.01]
Political Party	8	[-0.001, 0.02]	-0.11	[-0.21, -0.005]
Political Knowledge	-20	[-0.05, 0.08]	-0.34**	[-0.58, -0.11]
Overestimation Value	N/A	N/A	0.86	[-0.19, 1.9]

* indicates statistical significance at the 0.05 level, ** at the 0.01 level, and *** at the 0.001 level.

Experiment 3

Experiment 3 was intended to replicate and extend the findings of Experiment 2, by evaluating the impact of several specific design choices. First, the left-hand endpoint of the number line, 1 thousand in Experiments 1 and 2, was changed to the more natural value of zero. Second, the instructional manipulation used to exclude inattentive participants was made less difficult, and put before the rest of the experiment to eliminate the possibility of any selection bias that was dependent on condition. Third, a larger sample was used to better evaluate individual difference patterns.

The most theoretically substantive innovation in Experiment 3 was the inclusion of political opinion questions. In these questions, beyond asking what kind of impact a particular cut might have, we gave participants numerical information about the budget of two government programs (foreign aid, and the National Science Foundation), and asked whether they favored increasing or decreasing the funding for these programs. In the way people put marks on a line

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

predicts their quantitative judgments generally, and if quantitative judgments inform political views, then participants who see funding for a government program as proportionally small should tend to favor increasing it relative to those who see it as proportionally large.

Method

Participants.

A total of 968 participants American adults over the age of 18 were recruited in July of 2014 from MTurk after self-selecting the study, and were compensated 40 cents for 5-10 minutes of their time. 49 percent of participants identified as female and 77 percent as white, with an average age of 34. 53 percent of participants identified as a Strong Democrat, Weak Democrat, or Independent leaning Democrat, 21 percent identified as Independent, and 25 percent as a Strong or Weak Republican or Independent leaning Republican. As in Experiment 2, this sample is slightly younger and more liberal than the adult convenience samples reported by Berinsky et al. (2012).

Procedure

The format of Experiment 3 was very similar to Experiment 2: participants answered demographic and political knowledge questions, rated the impact of numerical and non-numerical political scenarios, completed a number line task, received the same training or control treatment, and completed a post-treatment number line task and rated more political scenarios. However, changes were made to the instructional manipulation check and the number line task, and political opinion questions were included as a final step in the procedure.

The instructional manipulation check was also changed and relocated among the demographic questions to prevent the elimination of participants who may have skimmed over

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

the instructions despite paying attention to the rest of the task, which was a potential concern in Experiment 2. Participants were asked, “Which job or office is currently held by Barack Obama?” but then before responding were asked to select one word that was mentioned in the previous sentence from a list of options. 17% of people failed this manipulation check and were eliminated from analysis, though the patterns that resulted were qualitatively identical when they were included. Because the manipulation check occurred before the treatment (which was the only variation between conditions), failure on this test could not have been influenced by treatment group.

While the number line task in Experiments 1 and 2 ranged from 1 thousand to 1 billion to discourage explicit calculations, which we presume voters do not make when reading news media, the number line in Experiment 3 ranges from zero to 1 billion. This change was made to ensure that the effects found in the previous experiments were not at all due to some special reasoning processes that take place with number lines starting at unusual values. Also, unlike Experiments 1 and 2, the post-treatment number line judgments were made on a slightly different set of numbers than the pre-treatment judgments, to avoid any possible practice effects (see online Appendix for specific stimuli).

Finally, in addition to asking participants to simply rate the impact of spending cuts on a greater budget or deficit, in Experiment 3 we explicitly assess how voters with higher levels of numeracy differ in their support for federal spending for two particular programs: NSF funding and foreign aid (Gilens 2001). After completing the post-treatment political scenario evaluations, following Gilens (2001) we asked participants to rate their support of current federal spending levels on two government programs. Participants were given short paragraph with current spending information, which included large numbers, for the National Science Foundation (NSF)

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

and foreign aid. One excerpt explained that, “in 2013, the U.S. federal government spent \$52 billion on foreign aid, including economic and military assistance to other countries. The total U.S. federal budget in 2013 was \$3.45 trillion”. Participants were then asked how much the U.S. should spend on foreign aid, with response options ranging from “a lot less” to “a lot more”.

Results & Discussion

A separate analysis of pre-treatment overestimation values revealed an unfortunate and statistically significant difference overestimation values before the treatment between the two scenario groups ($p \sim 0.008$): participants in the numerical group ($M(\text{overestimation value}) = 100$ million) happened to be more linear than participants in the word group ($M(\text{overestimation value}) = 140$ million). There was no statistical difference between the training and control groups (Training: $M(\text{overestimation value}) = 130$ million, Control: $M(\text{overestimation value}) = 114$ million) ($p \sim 0.16$). The only differentiating experience prior to this was random assignment. To reduce the impact of this difference, we decided to include pre-treatment overestimation value as a predictor in all models.

Number Line Placement

Number line placements were analyzed as in Experiment 1. Nonlinearity in number line placements was analyzed by a linear regression using pre-treatment overestimation values as a covariate, and both condition factors (treatment and scenario type) as independent measures. Statistical significance was estimated via a bootstrap with 10,000 random permutations to evaluate. Replicating Experiments 1 and 2, post-treatment overestimation values were significantly affected by the treatment type ($p = 0.001$; unstandardized $\beta = 50$ million; see Figure 8). There was no main effect of scenario type ($p = 0.80$), nor did scenario type interact with

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

overestimation value ($p=0.92$); this is expected, since the scenario type only varied for the political scenario evaluations. Pre-treatment overestimation value was also a substantial predictor of post-treatment overestimation value, suggesting at least some reliability to estimations ($p<0.0001$, unstandardized $\beta \sim 0.49$). In short, training significantly helped participants make more linear number judgments.

Political Scenario evaluations

As in Experiment 2, because raw scenario evaluation scores were unsurprisingly heavily skewed by the particular questions asked, raw scores were converted to z-scores and the resulting normalized scenario evaluations were analyzed.

Normalized scenario evaluations were submitted to an identical evaluation as number line placements, using pre-treatment scenario evaluations as a covariate. Because initial overestimation value differed between the scenario type conditions, and was predicted to have an impact on evaluations, initial overestimation value was also included in the analysis; however, excluding it did not change the character of the results. Once again, one-sided tests were used to evaluate predicted phenomena. There was a significant and moderate effect of treatment type on evaluations, ($p = 0.033$, $\beta = 0.27$). The interaction between the scenario type condition and the treatment did not reach statistical significance ($p \sim 0.16$); however, the effects of number line training were reliable when considering just participants who made numerical judgments ($p \sim 0.01$, Cohen's $d=0.27$), but not among participants who made non-numerical judgments ($p \sim 0.66$, Cohen's $d=0.04$). Thus, training in number line placement reduced impact evaluations when those impact estimates were numerical in character, and that training appears to be selective.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Because content differences between political scenarios made raw evaluation scores, for the main analysis (above), responses were normalized for each item, which allowed for direct comparison across treatments. Though the spending cuts featured in the political scenarios were larger than in Experiment 2, in proportional terms all cuts were still fairly small (less than 15%). However, as a check that people were at least somewhat sensitive to the numerical values used in the political scenarios, we evaluated, at pre- and post- test, whether larger numerical values were at least associated with larger impact estimates.

A bootstrapped linear regression was run on just the individuals who performed estimates on numerical stimuli, using actual proportions as predictors. Objectively larger cuts received marginally higher impact estimates before the treatment ($p \sim 0.07$) and significantly higher after ($p \sim 0.0000$). These results demonstrate that people are, in general, sensitive to the numerical magnitude of the cut whose impact they are estimating.

Political Opinions

The last two study questions related to political opinions, rather than judgments of “impact”—actual statements of support for increasing or decreasing the amount of US foreign aid and funding for the National Science Foundation. As a preliminary analysis, we submitted the judgment on each question to a repeated-measures analysis of variance, using party identification, pre-training overestimation value, specific question (NSF or foreign aid) and experimental condition (training or control) as independent predictors, as well as all interactions. This analysis revealed a significant main effect of pre-treatment overestimation value ($F(1,804)=7, p \sim 0.008$), such that people who more strongly overestimated small values on the number line were less supportive of government spending on these programs. However, the analysis also revealed a large and significant difference between the topics ($F(1, 804) \sim 1135$,

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

$p < 0.001$), as well as a significant interaction between overestimation value and topic ($f(1,804) \sim 8.6, p < 0.01$).

Because these questions are quite different in their popularity and relation to political ideology, and because the integrated analysis revealed differences between the two questions, we analyzed each separately. In both cases, we use pre-treatment overestimation value and condition (training vs. control) to predict support, so that we can evaluate the role of numerical cognition. In both cases, we allowed these factors to interact with self-reported party identification. To be clear: these are not intended as polling results; the population is self-selecting, and is certainly not representative. Our aim is to explore the existence of specific cognitive links between policies supported by some individual members of the electorate and their understanding of the numbers.

Figures 9 and 10 illustrate the major patterns. Self-identified Democrats in our sample were more supportive of increasing the National Science Foundation budget (64%) than were independents (55%) or Republicans (41%). This was significant by a bootstrap analysis ($p < 0.0001$). There was also a significant impact of overestimation value ($p \sim 0.01$): people who placed 1 million more than 5% of the way from the correct location were less supportive of expanding the nsf budget (46%) than were people with a more accurate line estimation of magnitudes involved (64%). Note that in this case the effect of overestimation is nearly as large as the gap associated with party identification in this sample. There was a marginally significant interaction between initial overestimation and party identification ($p \sim 0.07$), such overestimation tended to more strongly relate to judgments made by Democrats: Democrats who estimated ‘million’ as a smaller proportion of ‘billion’ (and therefore likely estimate ‘billion’ as a smaller portion of ‘trillion’) were substantially more supportive of increasing the budget of the NSF.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Interestingly, while number line judgments and a large influence on NSF support, and number line training affected assessments of political scenarios, number line training did not influence opinions about support for NSF (all p 's > 0.4).

More Republicans (74%) supported *cutting*⁵ foreign aid spending than did independents (63%) and Democrats (54%; $p < 0.0001$). In this case, there was not a significant relationship between overall overestimation and support for foreign aid spending ($p \sim 0.73$), but, the interaction between party and overestimation was statistically significant ($p \sim 0.028$); within Democrats, people with more linear estimates of 'million' were significantly ($p \sim 0.034$) more supportive of foreign aid spending levels (with 48% favoring cutting foreign aid) than were Democrats who overestimated small values (65%). Among Republicans, the reverse trend held, (favoring cutting fell from 75% among linear responders to 72% among non-linear responders) though the difference was not quite significant ($p \sim 0.060$).

Although statistical significance patterns shifted, across both questions a common pattern emerges: self-identified Democrats are sensitive to the perceived level of spending, being more supportive of increasing spending for programs when they perceive the current level to be small. In this sample, self-identified Republicans were relatively insensitive to perceived spending.

It should be emphasized that although there were robust relations between numeracy and opinions about these political programs, unlike scenarios evaluations, political opinions did not seem to depend on our short-term training treatment. There are two likely reasons for this

⁵ We switched from reporting active support for increasing NSF budget to supporting cutting foreign aid spending because the proportion of participants actively supporting increasing foreign aid spending was very low (8%). The cutoff of 5% for counting responses as "numerate" was also arbitrary, based on a natural break in the data. Analyses were conducted on the unbinned data, and so are unaffected by these variations

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

difference: first, it may be that participants form opinions on things like the NSF and foreign aid over time, and that the weak training conducted here was not powerful enough, or was not lasting enough, to overcome previous decisions reached using participants' baseline numeracy. Second, it is possible that it is not numeracy, per se, that is driving differences in viewpoint, but rather some individual differences factor that is correlated with our numeracy measure. As one check on this possibility, we reran all analyses using education, income, gender, age, and political knowledge as predictors of opinions, and found the same patterns reported here. Still, without an effect of an experimental treatment we cannot rule out the possibility that numeracy tracks, say, a generally pro-science or pro-technology attitude that drives differences in opinion.

Individual differences analyses

In order to evaluate how particular groups might differ in initial scenario evaluation, individual differences in pretest behavior were conducted. Two linear regressions using age, gender, political party identification, political knowledge, education level, and income as independent variables predicting numerical overestimation value in the first regression, and normalized scenario evaluation in the second (see Table 2). The entire sample was included for number line placement, but only the numeric judgment group was included for scenario evaluations. Since no strong predictions were made a priori, two-sided tests were used for all analyses regarding political affiliation, and Bonferroni correction was used in evaluating significance.

Overestimation was significantly impacted by political knowledge, such that participants with more political knowledge were also judged more numerate. Once again, males were more numerate than females. There were two trends: more educated people tended to be more

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

numerate than less educated people, and higher income participants tended to be more numerate.

However, neither of these trends was significant after Bonferroni correction.

Scenario evaluations were also affected by political knowledge, such that more knowledgeable participants gave lower estimated impact ratings, as did older participants. However, these patterns were dwarfed by the effect of pre-treatment overestimation, which strongly predicted scenario evaluations.

Table 2: Regression analysis of individual demographic factors in predicting numerical overestimation value and political scenario evaluation.

Variable	Overestimation Value		Scenario evaluations	
	β (in millions)	95% CI	β	95%CI
Education	-20*	[-0.04, -0.004]	0.1	[-0.22, 0.17]
Income	-8	[-0.02, -0.002]	0.02	[-0.06, 0.10]
Age	-1	[-0.002, 0.001]	-0.02*	[-0.04, -0.007]
Gender=Male	-68***	[-0.10, -0.04]	-0.27	[-0.58, 0.05]
Political Party	8	[-0.006, 0.022]	-0.05	[-0.14, 0.03]
Political Knowledge	-27**	[-0.04, -0.01]	-0.34**	[-0.54, -0.15]
Overestimation Value	N/A	N/A	0.49***	[0.80, 2.7]

* indicates statistical significance at the 0.05 level, ** at the 0.01 level, and *** at the 0.001 level.

General Discussion and Conclusion

Taken together, three experiments demonstrate that knowledge about large numbers is intertwined with political understanding in a broad though non-representative sample of the population.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

The results of Experiment 1 indicate that a very short training treatment involving a single piece of information—the correct placement of a single number on a number line—affects subsequent number-line estimations. This sort of one-shot training is commonly seen in small number number-line tasks (Thompson and Opfer 2010; Siegler and Opfer 2003); this the first time it has been demonstrated for large number tasks. The logarithmic or compressive patterns seen in young children are widely shared across animals species and seem to reflect early-developing biases in the comprehension of magnitude, while the errors seen in large-number tasks are rooted at least in part in the specific notation used. Nevertheless, one-shot training paradigms seem to be effective in both cases.

Experiment 2 provides new results in the psychology of political science that also have the potential for practical applicability. In particular, Experiment 2 shows that exactly the same kind of very short, one-shot, and entirely non-political training selectively, substantially, and quickly affects the judgments people make about real-world United States political scenarios—at least for judgments made immediately after training. It is plausible on the basis of these results that a substantial portion of the population would react quite differently to political actions related to budgetary decisions if they had a clear understanding of the numerical values involved. Moreover, these results suggest that a quick non-political training could facilitate such a clear understanding of numerical values.

These results suggest that a common core of knowledge about number representations and magnitudes is invoked in both generic numerical tasks and in political reasoning. This contrasts with other contexts, in which quantitative reasoning has been shown to be either heavily biased by politics (Kahan et al. 2013), or entirely undetectable (Lawrence and Sides 2011; Sides and Citrin 2007). As demonstrated in Experiment 3, political opinions themselves

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

differ across participants of different numeracy, and in the direction that a quantitative reasoning model would predict: people support government programs more when they perceive the programs to be relatively cheap, compared to when they misperceive the relative costs. However, this numerical sensitivity is larger among self-identified Democrats than Republicans; it may be that Republicans are more likely to oppose on principle the particular programs we tested, without regard for the cost.

We conclude from these results that number knowledge is a component of political understanding and opinion-formation in at least one substantial voting-age population (users of Mechanical Turk). The use of a nationally representative sampling technique would be required to estimate the exact percentage of the population and subpopulations exhibiting this behavior, and to make strong claims about, for instance, how numeracy training is likely to impact broad political support for programs such as funding for scientific research or foreign aid. We do not expect the results reported here to be representative of the population; rather, they demonstrate that (and clarify how) number perception plays a mechanistic role in the political opinions of at least some US voters. Obviously many other mechanisms also play large roles.

In addition to the short time required and the simplicity of the kind of numerical training used in these Experiments, one appeal is its generalizability. Rather than providing separate additional information to voters on such issue-specific contexts as the budget for the legislative branch of government, we used a single example to change participants' views on scenarios involving each of these topics. In a related line of research, some of the authors (Guay et al, under review) report long-term changes to political perception in a very specialized and motivated population after a somewhat more detailed instructional treatment. In particular, they found that a five-minute lesson on the place-value system provided in the context of an

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

introductory political science course shifted how students evaluated budget scenarios 8 weeks after the lesson. Putting together those results with the patterns reported here, there is at least some reason to believe that the political views of the general population are colored by large-number innumeracy, and that effective training treatments can be developed that mitigate that innumeracy. Further research is necessary to determine whether training like that provided in these studies has a long-term effect on the general-population, and whether such a treatment would have practical applicability to improve decision-making by voters and politicians on the pervasive variety of modern issues that involve the use of large numbers. Furthermore, the distribution of error-prone reasoning is largely unknown. For instance, it is currently unclear whether politicians and other individuals involved in creating budgets are themselves prone to misinterpretation of the relevant estimates, or whether this error is restricted to people outside government and without great wealth, for whom these numbers are perhaps primarily encountered in political journalism and speeches.

What we can see is that errors in reasoning about large numbers are widespread, are predictable in character, impact judgments, and can be ameliorated through education. Given the central role of the kinds of numbers involved in this study to current national and international debates about austerity, investment, and even government default, understanding how to predict and repair these errors has never been more important.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Online Appendix

Number Line Estimations

Table 3: Specific numerical stimuli used in Experiments 1, 2, and 3.

Experiment	Numerical Stimulus Values
1, 2, and 3-Post	260 Thousand; 470 Thousand; 670 Thousand ; 920 Thousand; 3 Million; 280 Million; 530 Million; 840 Million
3-Pre	320 Thousand, 430 Thousand,, 760 Thousand,, 870 Thousand,, 5 Million, 340 Million, 480 Million, 810 Million

Political Scenario Evaluation Questions

Response Options

The following response options were presented after each political judgment question:

No impact, A very small impact, A small impact, A somewhat small impact, A somewhat large impact, A large impact, A very large impact, A huge impact

Numerical Condition Questions

"The non-partisan commission advised that through executive action, congressional rule, and legislation, a number of steps be taken immediately to show Washington can lead by example. These collected recommendations would reduce spending on both the security and non-security sides of the discretionary budget. Together, they will save more than \$50 billion in 2012 alone". [The projected deficit for 2012 is \$1.33 trillion]

What was the impact of the spending reduction on the projected 2012 deficit?

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

"Congress has emphasized that nearly every corner of the government is on the chopping block, and their newly passed proposal includes a reduction of \$194 million in spending for the legislative branch". [The legislative branch's budget in 2011 was \$5.54 billion]

What was the impact of this proposal on the legislative branch's 2011 budget?

"Republican lawmakers took the hatchet to funding for their own legislative branch with a cut of \$103 million". [The legislative branch's budget in 2011 was \$5.54 billion]

What was the impact of these cuts on the legislative branch's 2011 budget?

"Congressmen on the House Appropriations Committee took what its chairman called a 'wire brush' to the budget, proposing the removal of \$58 billion in non-security discretionary budget funds". [The federal budget deficit in 2011 was \$1.33 trillion]

What was the impact of the removal of the non-security discretionary funds on the 2011 federal deficit?

"According to the plan, the federal Institute of Museum and Library Services would see a \$16 million cut under the stopgap measure. That's the amount of money the agency received in earmarks to help museums and libraries across the nation build exhibits, buy supplies and fund other projects last year". [House leaders are currently trying to cut \$30 billion from the federal budget]

What was the impact of the spending cut on the \$30 billion spending-cut goal?

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

"The House vote came just before dawn on Saturday, capping an extraordinary week-long debate that featured many amendments and a bill that cut \$61 billion from current spending levels". [The projected deficit for 2012 is \$1.33 trillion]

What was the impact of the spending reduction on the projected 2012 deficit?

"A government official proposed \$17 billion in spending cuts today, including cuts in weapons systems already announced, according to two senior staff members familiar with the plans". [The projected deficit for 2012 is \$1.33 trillion]

What was the impact of the spending cuts on the projected 2012 deficit?

"This plan includes many spending cuts; U.S. Customs and Border Protection, for example, must trim \$512 million from their budget". [The U.S. Customs and Border Protection budget in 2011 was \$11.5 billion]

What was the impact of the spending reduction on the U.S. Customs and Border Protection 2011 budget?

Word Condition Questions

"Yesterday, the government announced a plan that will give funds to U.S. auto companies. The plan gives money this month and next month to General Motors Corporation and Chrysler LLC. Both companies have said that they soon might be unable to pay their bills without federal help".

What was the impact of the U.S. government's funding of the auto companies on the economy?

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

"Barack Obama says the 39th anniversary of the Roe v. Wade Supreme Court Decision on abortion is the chance to recognize the 'fundamental constitutional right' to abortion and to 'continue our efforts to ensure that our daughters have the same rights, freedoms, and opportunities as our sons to fulfill their dreams'".

What was the impact of Obama's statement on the national abortion debate?

"With final action in the Senate today, Congress has approved legislation that formally recognizes the U.S. obligation to withdraw from Iraq and requires the Pentagon to provide quarterly reports on the progress of that withdrawal".

What was the impact of the newly passed legislation on the U.S. war on terror?

"President John Fitzgerald Kennedy was shot and killed by an assassin yesterday. He died of a wound caused by a rifle bullet that was fired at him as he was riding through downtown Dallas in a motorcade".

What was the impact of the president's assassination on the stability of the U.S.?

"After two weeks of anguishing debate, Congress passed the plan to save the financial industry and prop up the economy by giving funds to U.S. banks".

What was the impact of the government's funding of U.S. banks on the economy?

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

"On Monday, Barack Obama abandoned plans for a civilian trial in New York City for the self-proclaimed mastermind of the Sept. 11 attacks and is sending the case to a military commission at Guantanamo Bay, Cuba".

What was the impact of Barack Obama's decision on the U.S. war on terror?

"The U.S. House of Representatives impeached President Clinton yesterday on charges of obstructing justice and lying to a federal grand jury about his sexual affair with a former White House intern".

What was the impact of the President's impeachment on the stability of the United States?

"In an interview with ABC News, Barack Obama declared his support for gay marriage. This marks a departure from the president's previous stance, which has repeatedly been described as 'evolving'".

What was the impact of Barack Obama's support for gay marriage on the gay marriage debate in the U.S.?

Unpublished Draft!

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Acknowledgments

This research was partially supported by Department of Education, Institute of Education Sciences grant R305A110060. We would like to thank the members of the Cave Lab at the University of Richmond, Erin Ottmar, and Ernest McGowen for their contributions and feedback. The data reported in Experiment 1 were included as Experiment 2 of a 6-page report and poster presented at the Proceedings of the Cognitive Science Society.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

References

- Barth, H.C., & Paladino, A.M. (2011). The development of numerical estimation: evidence against a representational shift. *Developmental Science*, 14, 125–135.
- Barth, H., Slusser, E., Cohen, D., & Paladino, A. (2011). A sense of proportion: commentary on Opfer, Siegler, and Young. *Developmental Science*, 14(5), 1205-1206.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). Linear mixed-effects models using Eigen and S4 (Version 1.0-5) [R package]. Available from <http://CRAN.R-project.org/package=lme4>
<http://cran.r-project.org/package=lme4>
- Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experiment research: Amazon.com. *Political Analysis*.
- Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. Translated by Louise Sommer. *Econometrica* 22(1), 23-36.
- Booth, J.L., & Siegler, R.S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 42, 189-201.
- Buhrmester, M., Kwang, T., & Gosling, S.D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science*, 6, 3-5.
- Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic bulletin & review*, 18(2), 331-338.
- Crawford, J. T., & Pilanski, J. M. (2012). Political intolerance, right and left. *Political Psychology*.
- Dehaene, S., & Brannon, E. (Eds.). (2011). *Space, time and number in the brain: Searching for the foundations of mathematical thought*. Academic Press.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Delli Carpini, M.X., & Keeter, S. (1996). What Americans know about politics and why it matters. Yale University Press.

Druckman, J.N. (2004). Priming the vote: Campaign effects in a U.S. Senate election. *Political Psychology*, 25, 577-594.

Fausey, C. M., & Matlock, T. (2011). Can grammar win elections?. *Political Psychology*, 32, 563-574.

Fazio, L.K., Bailey, D.H., Thompson, C.A., & Sielger, R.S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, 123, 53-72.

Fiske, S.T., Kinder, D.R., & Larter, W.M. (1983). The novice and the expert: Knowledge-based strategies in political cognition. *Journal of Experimental Social Psychology*, 19, 381-400.

Gerber, A.S., Huber, G.A., Doherty, D., Dowling, C.M., & Panagopoulos, C. (2013). Big five personality traits and responses to persuasive appeals: Results from voter turnout experiments. *Political Behavior*, 35, 687-728.

Gilens, M. (2001). Political ignorance and collective policy preferences. *The American Political Science Review*, 95, 379-396.

Guay, B., Chandler, C., Erkulwater, J., & Landy, D. (under review). Testing the Effectiveness of a Number-based Classroom Exercise.

Henry, P. J. (2008). College sophomore in the laboratory redux: Influence of a narrow data base on social psychology's view of the nature of prejudice. *Psychological Inquiry*, 19, 49-71.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Iyengar, S., & Kinder, D.R. (1987). News that matters: Agenda setting and priming in a television age. *News that matters: Agenda setting and priming in a television age*.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2013). Motivated Numeracy and Enlightened Self-Government. Available at SSRN: <http://ssrn.com/abstract=2319992>

Krupnikov, Y., & Bauer, N.M. (forthcoming). The relationship between campaign negativity, gender and campaign context. *Political Behavior*.

Kuklinski, J. F., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R. F. (2000). Misinformation and the currency of democratic citizenship. *The Journal of Politics*, 62, 790-816.

Landy, D., Charlesworth, A. & Ottmar, E. (2014). Cutting in line: discontinuities in the use of large numbers in adults. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Quebec City, Quebec: Cognitive Science Society.

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive Science*, 37, 775–799.

Landy, D., Silbert, N., & Goldin, A. (2012). Getting Off at the End of the Line: The Estimation of Large Numbers. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*.

Lawrence, E.D., & Sides, J. (2011). The consequences of political innumeracy. Unpublished paper presented at Annual Meeting of the American Political Science Association, Seattle, WA.

Link, T., Huber, S., Nuerk, H.C., & Moeller, K. (2013). Unbounding the Mental Number Line- New Evidence on Children's Spatial Representation of Numbers. *Front. Psychology* 4: 1021.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Lodge, M., & Hamill, R. (1986). A partisan schema for political information processing. *The American Political Science Review*, 80, 505-520.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44, 1-23.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32, 303-330.

Opfer, J. E., Siegler, R. S., & Young, C. J. (2011). The powers of noise-fitting: Reply to Barth and Paladino. *Developmental Science*, 14, 1194-1204.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45, 867-872.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5, 411-419.

Rahn, W.M. (1993). The role of partisan stereotypes in information processing about political candidates. *Midwest Political Science Association*, 37, 472-496.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child development*, 79(2), 375-394.

Representative Fattah (PA). "Commerce, Justice, Science, and Related Agencies Appropriations Act, 2013." *Congressional Record* (May 9, 2012) p. H2544.

Resnick, I., Shipley, T.F., Newcombe, N., Massey, C., & Wills, T. (2012). Examining the Representation and Understanding of Large Magnitudes Using the Hierarchical Alignment

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Model of Analogical Reasoning. Proceedings of the 34th Annual Conference of the Cognitive Science Society.

Rips, L.J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of a number. *Behavioral and Brain Sciences*, 31, 623-687.

Schaffner, B.F. (2011). Racial salience and the Obama vote. *Political Psychology*, 32, 963-988.

Sides, J. & Citrin, J. (2007). How large the huddled masses? The causes and consequences of public misperceptions about immigrant populations. Unpublished paper presented at Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Siegler, R.S., & Booth, J.L. (2004). Development of numerical estimation in young children. *Child Development*, 75, 428-444.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, 14, 237–243.

Siegler, R. S., & Robinson, M. (1982). The development of numerical understanding. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior*, Vol. 16 (pp. 242–312). New York: Academic Press.

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 683-692.

Stevens, S. S. (1975). *Psychophysics*. Transaction Publishers.

Taber, C. (2003). Information processing and public opinion. In D. O. Sears, L. Huddy & R. Jervis (Eds.), *Handbook of political psychology*. London, England: Oxford University Press.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Thompson, C.A., & Opfer, J.E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development*, 81, 1768-1786.

Travers, M. W., Van Boven, L., & Judd, C. J. (2013). The secrecy heuristic: Inferring quality from secrecy in foreign policy contexts. *Political Psychology*.

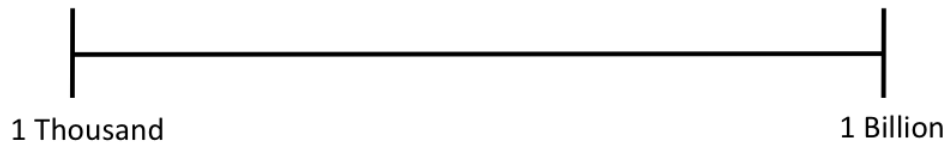
Tversky A., Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297-323.

Utah data center. (2013). Retrieved from <http://nsa.gov1.info/utah-data-center/>.

Visser, P.S., Krosnick, J.A., & Lavrakas, P. (2000). Survey research. In H.T. Reis & C.M. Judd (Eds.), *Handbook of research methods in social psychology*. New York: Cambridge University Press.

WorldPublicOpinion. (2010) American Public Opinion on Foreign Aid. Available at http://www.worldpublicopinion.org/pipa/pdf/nov10/ForeignAid_Nov10_quaire.pdf.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS



280 million: Please choose a place on this number line that corresponds to the value 280 million. Mark your choice by clicking on the line. You may not be able to see the mark you place—That's okay! We'll record it. If click multiple times, we'll get your last click.

Figure 1: The number line task instructs participants to place a number (280 million) on a horizontal line with two numerical endpoints (here, 1 thousand and 1 billion)

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

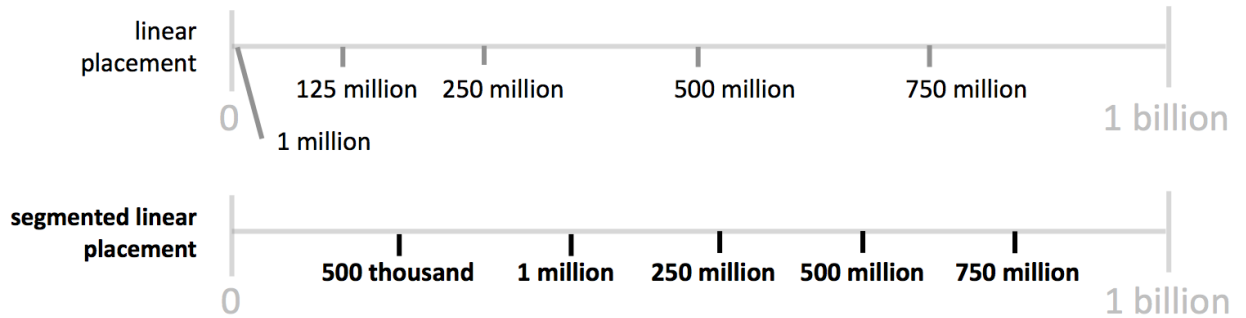


Figure 2: The top number line shows the correct linear number line placements, while the bottom number line demonstrates innumerate “hockey stick” number line placements, which about 40% of participants made. The numerate group placed numbers in the thousands range linearly between 0 and 1 million, but overestimated the magnitude of 1 million, causing drastic overestimations.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

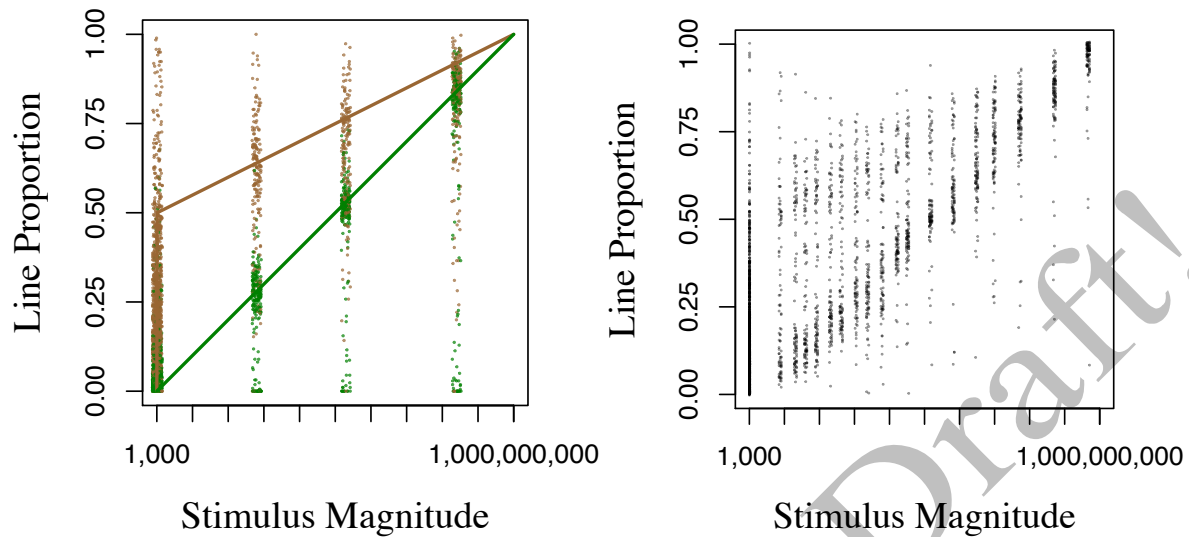


Figure 3: Raw participant responses on a number line task, with the numerical value (x axis) plotted against the line proportion. The two typical patterns of response can clearly be seen: The numerate response is a straight line from the lower left to the upper right, while the ‘hockey stick’ pattern shows as a straight vertical line, bending at around the halfway mark to make a second straight line to the upper right. This corresponds to a million overestimate of 500 million. The left panel is raw data from the current experiment 1. In this set, the x axis has been jittered slightly to better show the data points, and participants have been binned by a median split on; the subjective value of 1 million (the overestimation value); the right panel is data originally reported in Landy, Silbert, and Goldin, 2013, Figure 2, panel A, showing the same hockey stick pattern in a larger set of stimuli.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

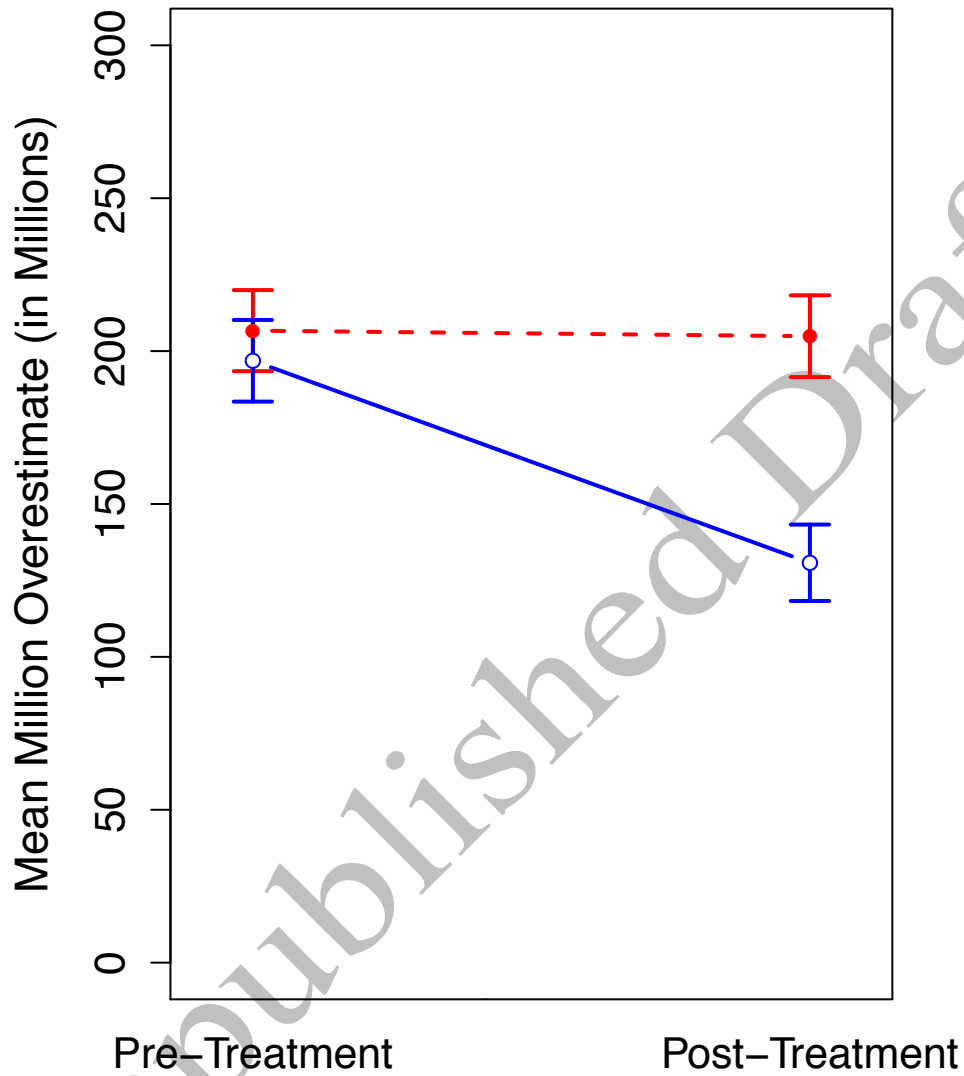


Figure 4: Mean estimated value of 1 million (overestimation value) in Experiment 1

before and after treatment. Points represent mean values; errors are standard errors.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

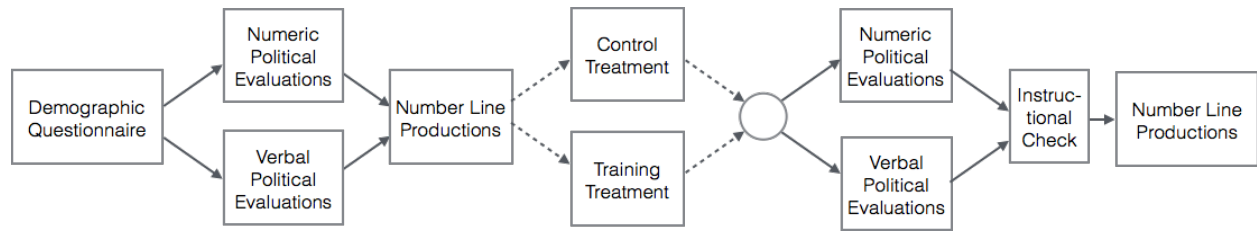


Figure 5: Design of Experiment 2. The dashed and solid arrows reflect independent between-subjects factors. Each participant was separately assigned to a Treatment and to a judgment type. Participants failing the instructional manipulation check were eliminated from the analysis. The circle reflects that the two factors were crossed; i.e., that participants from both control and training groups were randomly assigned to numeric or word scenario evaluations.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

Figure 6: Mean estimated location of 1 million (*overestimation value*) in Experiment 2 before and after Treatment. Points represent mean values; errors are standard errors. As in Experiment 1, participants receiving the training treatment increased accuracy on the number line task more than did those in the control treatment.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

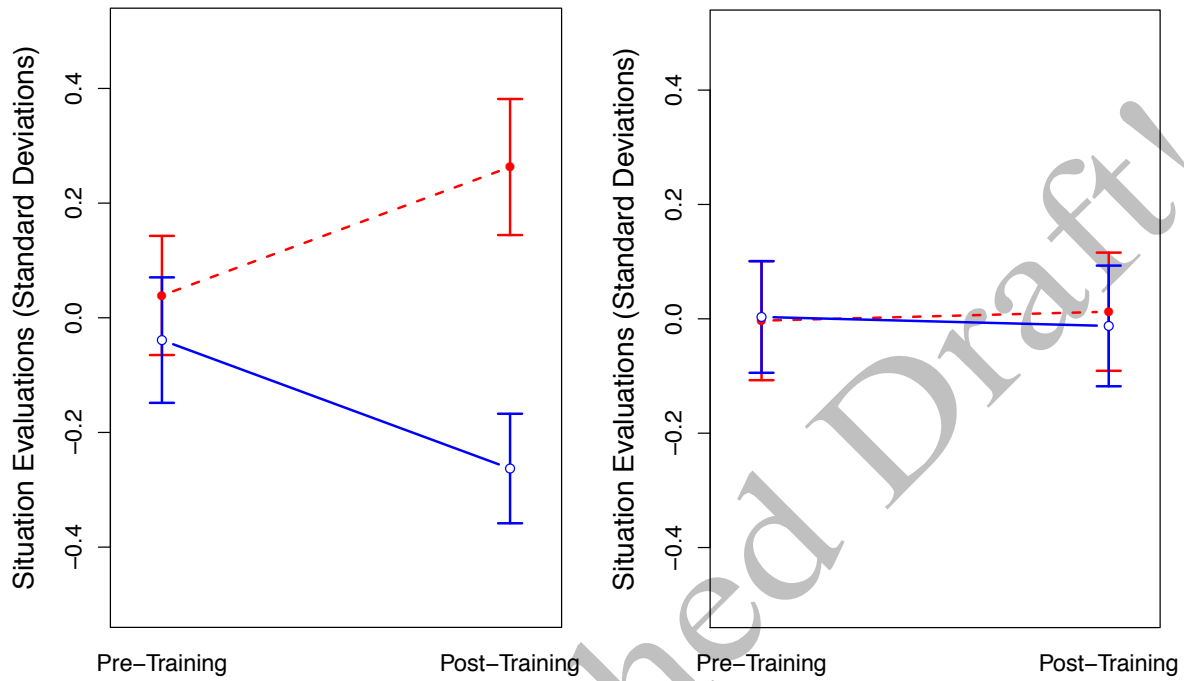


Figure 7: Mean estimated impact for political scenario evaluations in Experiment 2 before and after treatment in the numerical (left) and word (right) evaluation condition. Points represent individual participant evaluations by condition. Errors are standard errors

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

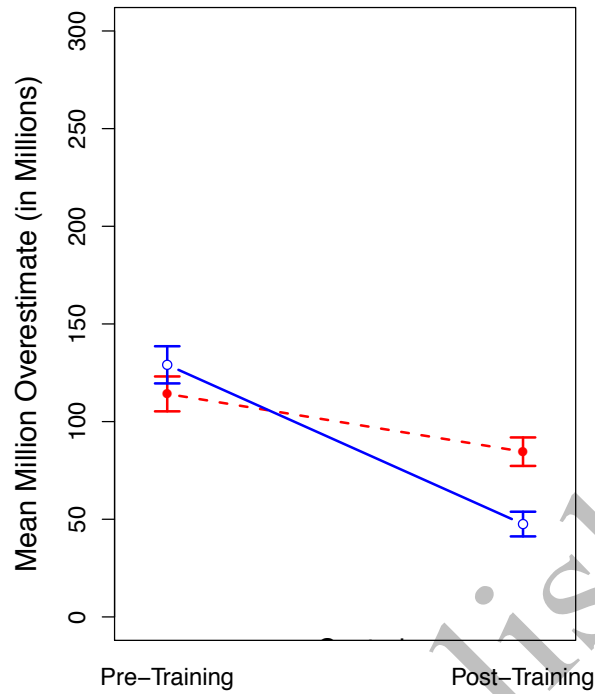


Figure 8: Mean estimated location of 1 million (*overestimation value*) in Experiment 3 before and after treatment. Points represent mean values; errors are standard errors. As in Experiment 1, participants receiving the Training treatment increased accuracy on the number line task more than did those in the control treatment.

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

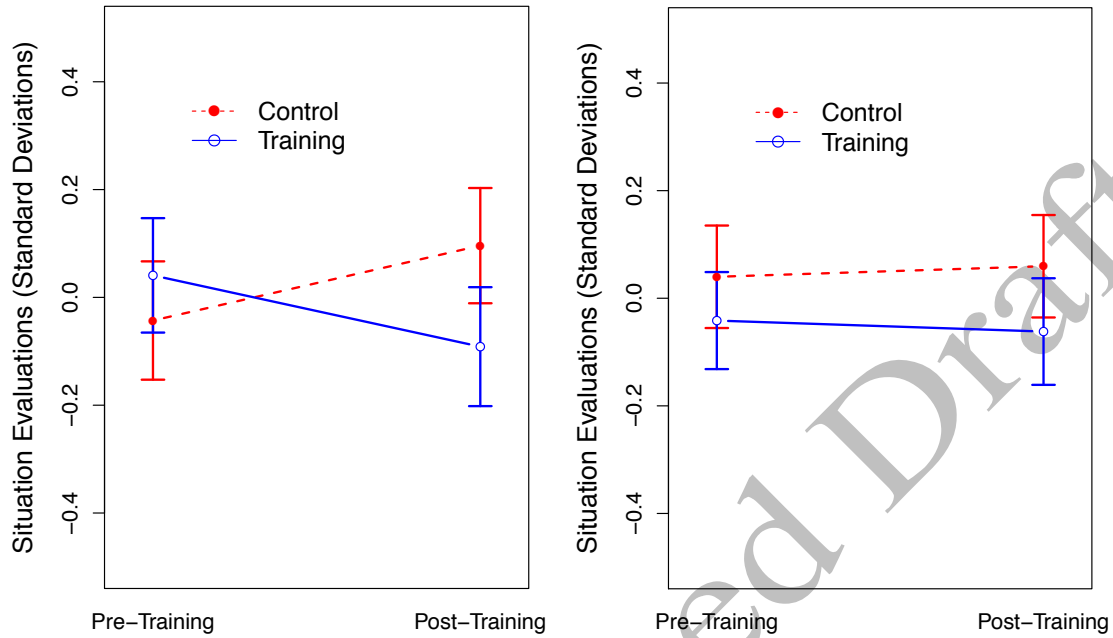


Figure 9: Mean estimated impact for political scenario evaluations in Experiment 3 before and after treatment in the numerical (left) and word (right) evaluation condition. Points represent individual participant evaluations by condition. Errors are standard errors

NUMBER COMPREHENSION IMPACTS POLITICAL JUDGMENTS

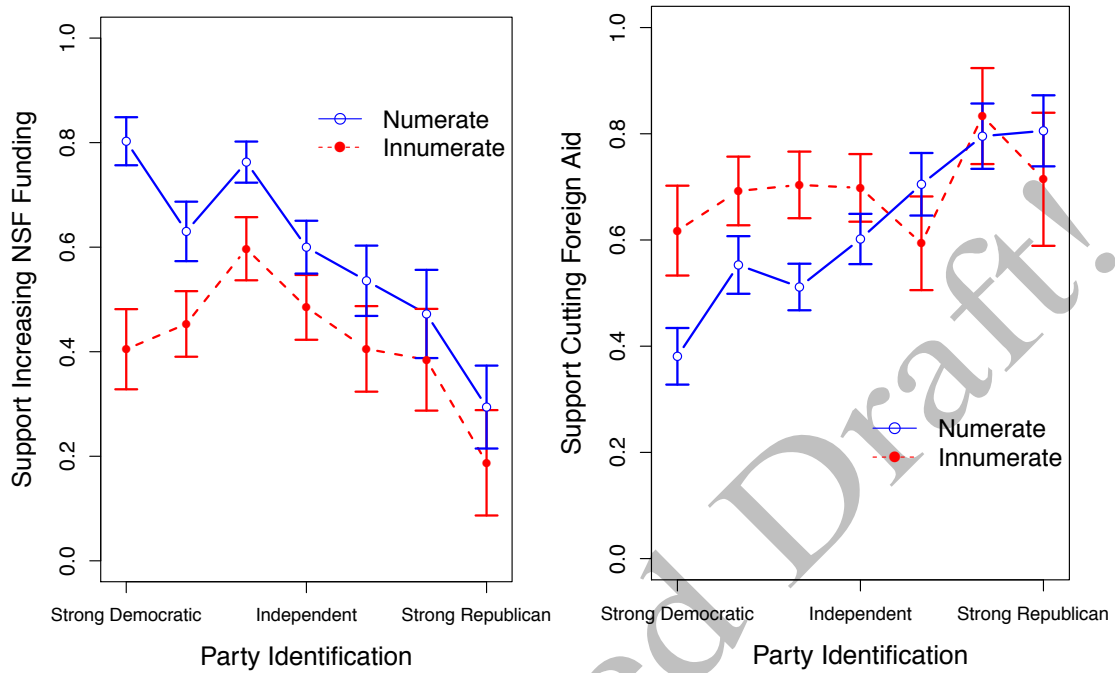


Figure 10: Expressed support for increasing NSF funding, and cutting foreign aid in Experiment 3, separated by party identification and numeracy category. Participants were treated as “numerate” if their subjective value of 1 million (the overestimation value) was within 5% of the correct answer. Participants expressed preferred change in each program’s funding level on a 5-point Likert scale ranging from “A lot less” to “a lot more”, summarized in this diagram by a binary ‘support’ variable for illustrative purposes.