Published in final edited form as:
J Behav Decis Mak. 2012 October ; 25(4): 361–381. doi:10.1002/bdm.752.

Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment

JORDANA M. LIBERALI[1,4], **VALERIE F. REYNA**[2,*], **SARAH FURLAN**[3], **LILIAN M. STEIN**[4], and **SETH T. PARDO**[2]

[1]Erasmus University Rotterdam, Rotterdam, The Netherlands [2]Cornell University, Ithaca, NY, USA [3]University of Padova, Padova, Italy [4]Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil

Abstract

Despite evidence that individual differences in numeracy affect judgment and decision making, the precise mechanisms underlying how such differences produce biases and fallacies remain unclear. Numeracy scales have been developed without sufficient theoretical grounding, and their relation to other cognitive tasks that assess numerical reasoning, such as the Cognitive Reflection Test (CRT), has been debated. In studies conducted in Brazil and in the USA, we administered an objective Numeracy Scale (NS), Subjective Numeracy Scale (SNS), and the CRT to assess whether they measured similar constructs. The Rational–Experiential Inventory, inhibition (go/no-go task), and intelligence were also investigated. By examining factor solutions along with frequent errors for questions that loaded on each factor, we characterized different types of processing captured by different items on these scales. We also tested the predictive power of these factors to account for biases and fallacies in probability judgments. In the first study, 259 Brazilian undergraduates were tested on the conjunction and disjunction fallacies. In the second study, 190 American undergraduates responded to a ratio-bias task. Across the different samples, the results were remarkably similar. The results indicated that the CRT is not just another numeracy scale, that objective and subjective numeracy scales do not measure an identical construct, and that different aspects of numeracy predict different biases and fallacies. Dimensions of numeracy included computational skills such as multiplying, proportional reasoning, mindless or verbatim matching, metacognitive monitoring, and understanding the gist of relative magnitude, consistent with dual-process theories such as fuzzy-trace theory.

Keywords

numeracy; fuzzy-trace theory; cognitive reflection; ratio bias; conjunction fallacy; disjunction fallacy; intelligence

---

There has been increasing attention paid to individual differences in judgment-and-decision-making research over the past decade, including research on developmental differences that can be thought of as a type of individual difference (Weber & Johnson, 2009). Such individual differences have implications for the real world because they imply that some people are likely to make better medical, legal, or policy decisions than others; identifying

*Correspondence to: Valerie F. Reyna, Cornell University, B44 Martha Van Rensselaer Hall, Ithaca, NY, USA., vr53@cornell.edu.

these individuals has the potential to improve outcomes for the broader society (Nelson, Reyna, Fagerlin, Lipkus, & Peters, 2008; Reyna & Farley, 2006). Further, research on individual differences has been used to adjudicate important theoretical controversies, especially regarding biases and fallacies in judgment and decision making (e.g., Evans, 2007; Milkman, Chugh, & Bazerman, 2009; Stanovich & West, 2000, and commentaries).

Despite the proliferation of new scales and tasks, uncertainty remains about exactly what they are measuring and how they relate to biases and fallacies. Among these scales and tasks, numeracy scales and the Cognitive Reflection Test (CRT) seem to hold particular promise for understanding and predicting behavior (Frederick, 2005; Reyna, Nelson, Han, & Dieckmann, 2009). Numeracy refers to the ability to understand and use numbers, and it has been shown to be important in a range of everyday tasks, such as medical decision making. Low numeracy also increases susceptibility to a variety of biases and fallacies, even when general intelligence is partialled out (e.g., Peters et al., 2006). Similarly, the CRT ("cognitive reflection" is defined as "the ability or disposition to resist reporting the response that first comes to mind," p. 35, Frederick, 2005) has been shown to predict time and risk preferences, including preferences for options with higher expected value and resistance to logical fallacies (Campitelli & Labollita, 2010; Cokely & Kelley, 2009; Oechssler, Roider, & Schmitz, 2009).

As Reyna et al. (2009) argue, despite their usefulness, a limitation of the measures of numeracy is that they are not theoretically motivated. Frederick (2005), too, points out that the CRT shares features with other tests of cognitive ability and style (e.g., Need for Cognition), but the cognitive processes being assessed are not well understood or precisely distinguishable from processes tapped by other measures. Indeed, performance on numeracy tests and on the CRT has been found to correlate positively (Cokely & Kelley, 2009), and some have speculated that these quantitative tasks measure similar constructs.

In this article, we empirically examine the question of what these scales and tasks measure by applying factor analysis to items on objective numeracy scales, subjective numeracy scales, and the CRT and by interpreting the factors in terms of modern cognitive theories. Our interpretation was aided not only by examining commonalities among items that load together but also by examining frequent answers participants gave for each item. Using regression, we then related these theoretically interpreted factors to biases and fallacies in probability judgment (Tversky & Kahneman, 1974) to more deeply understand the mechanisms that might account for previously observed correlations between numeracy and the CRT, on the one hand, and these biases and fallacies, on the other hand. Thus, the main issues we consider include whether CRT is another numeracy scale, whether objective and subjective numeracy scales measure the same construct, and how factors that characterize these scales predict biases and fallacies in probability judgment.

Specifically, we conducted two studies, one with college students in Brazil and the other with college students in the USA, that each assessed objective numeracy, subjective numeracy, and the CRT. We conducted systematic factor analyses on these core measures in each dataset to assess the degree to which they measure similar or different constructs. On the basis of prior literature in mathematical cognition, we expect that some factors should reflect conceptual knowledge, such as linear representations of relative magnitude (i.e., sometimes called ordinal gist) or understanding of proportions or ratios, whereas other items should reflect procedural knowledge of mathematical operations, such as multiplication (Bouwmeester, Vermunt, & Sijtsma, 2007; National Mathematics Panel, 2008; Reyna, 2008; Reyna & Brainerd, 1994, 2007; Siegler & Opfer, 2003). Building on early distinctions between unthinking (mindless) and meaningful reasoning in Gestalt theory, fuzzy-trace theory also predicts that lower level reasoning can be characterized by "verbatim" responses that match elements of questions, in contrast to gist-based responses that go beyond surface elements (a matter of degree because reasoning varies along a verbatim-to-gist continuum; see Brainerd & Reyna, 1992; Kahneman, 2003; Reyna, Lloyd, & Brainerd, 2003). Still, other items, such as those on SNS and CRT, have been hypothesized to reflect metacognitive judgment or monitoring in which initial (wrong) answers may be censored (Dunning, Heath, & Suls, 2004; Frederick, 2005; Stanovich & West, 2008).

Further, in both studies, we assessed the predictive validity of the obtained factors in accounting for biases and fallacies. In the first study, we assessed conjunction and disjunction fallacies (Reyna & Mills, 2007a; Tversky & Kahneman, 1983). A conjunction fallacy is judging the joint probability of events (one event *and* another event occurring) to be more likely than one of the component events. A disjunction fallacy is judging the probability of the disjunction of events (one event *or* another event occurring) as less likely than one of the component events. Because of theoretically predicted relations between distortions in memory and in probability judgments (e.g., Reyna & Kiernan, 1994; Wolfe & Reyna, 2010), we also assessed participants' memories for the frequencies of events that were the basis of the conjunctive and disjunctive probability judgments. In the second study, we assessed the ratio bias in probability judgment (e.g., preferring 10/100 chances to win over 1/10 chances to win, despite their numerical equivalence). Although studies have linked overall numeracy and the CRT to these biases and fallacies (see Reyna et al., 2009, for a review), we explore the predictive validity of factors underlying these scales.

In addition to the core measures of numeracy and CRT, we expanded our measures of individual differences in the second study to include Need for Cognition, Faith in Intuition, inhibitory control (the go/no-go task), and general intelligence. According to most accounts, individuals who score high in numeracy should be less likely to exhibit biases and fallacies because they think more rationally and objectively, so-called Type 2 thinking (Epstein, 1994; Epstein, Pacini, Denes-Raj, & Heier, 1996; Lipkus & Peters, 2009; Peters et al., 2006; Stanovich & West, 2008). Therefore, in the second study, we added a measure of rational or Type 2 thinking (Need for Cognition) and of intuitive or Type 1 thinking (Faith in Intuition), the two dimensions of the Rational–Experiential Inventory (REI, Pacini & Epstein, 1999). The REI was designed specifically to predict the ratio bias (and was extended to other biases) and has also been linked to explanations of numeracy (Reyna & Brainerd, 2008). The go/no-go task, a measure of inhibition, was added to determine whether performance on the core measures of numeracy and the CRT could be explained in part by an ability to inhibit (i.e., censor) intuitive responses (e.g., Frederick, 2005; Kahneman, 2003). Finally, a measure of intelligence was added to rule out an alternative explanation for what numeracy scales and CRT measure and why they predict biases and fallacies, namely that they are merely measures of intelligence (but see Peters et al., 2006, for evidence against this hypothesis).

STUDY 1: BRAZILIAN SAMPLE

Method

Participants—Participants were 259 undergraduate students (mean age= 24.04years) from three different courses (210 from management, 32 from engineering, and 17 from accounting) of three Brazilian universities (115 women, 142 men, and 2 did not specify). It was a convenience sample, gathered through a snowball technique applied to professors and lecturers. All participants gave written informed consent, and the study was approved by the Institutional Review Board of *Pontifícia Universidade Católica do Rio Grande do Sul*.

Materials and procedure—Students participated in a probability learning (also called experiential learning) paradigm similar to the Iowa gambling task (Bechara, Damasio, Tranel, & Damasio, 1997). The participants were presented sequentially with 20 dinners in a random order that each of two fictitious characters had last month (40 dinners total), presented one at a time in 40 slides; in each slide, a photo of the face of the character accompanied a phrase describing the meal (e.g., Álvaro had grilled chicken). Target frequency was manipulated so that each character was associated with a high frequency target (presented 12 times), a medium–high frequency target (presented five times), a medium–low frequency target (presented two times), and a low frequency target (presented only one time). The "gist" of the meals for one character was a clear preference for red meat (unhealthy) and for the other character was chicken and fish (healthy). The gist was not explicitly presented.

After viewing the target material, the participants were asked to estimate the probability that each person would have had a given meal (or combination of two meals, forming conjunctions or disjunctions) for dinner last month (past judgments) or next month (future judgments) (e.g., What is the probability that Cristiano will have top sirloin for dinner next month?). Past and future judgments were blocked, and the order of blocks was counterbalanced across the participants. Probability judgments were made about single meals (e.g., What is the probability that Álvaro will have poached fish for dinner next month?), conjunctions of meals (e.g., What is the probability that Cristiano will have top sirloin *and* prime rib for dinner next month?), and disjunctions of meals (e.g., What is the probability that Cristiano will have top sirloin *or* prime rib [or both] for dinner next month?). To convey their probability judgments, the participants selected a number from 0% (described as "impossible") to 100% (described as "absolutely certain"); 50% was described as meaning "as likely as not."

Memory for presented meals (which varied in frequency) was assessed using a cued recall test. The participants were asked how many times each meal was presented (e.g., Out of 20 meals, how many times did Álvaro eat grilled chicken?) for targets (presented meals), related items (meals that were never presented but were consistent with the gist of the meals that were presented), and unrelated items (neither presented nor gist consistent). The correct frequency estimates for the related and unrelated items is zero. Each participant received a memory deviation score (i.e., the degree to which remembered frequencies of events differed from presented frequencies) and a gist memory score (i.e., the summation of all related distractors' frequency judgments minus the summation of all unrelated distractors' frequency judgments).

We assessed individual differences through the following measures:

Numeracy scales: Objective numeracy tests contain items that assess basic probability and mathematical concepts including simple mathematical operations on risk magnitudes using percentages and proportions, as well as conversion of percentages to proportions, and vice versa (Reyna & Brainerd, 2007; Table 1). In this study, the participants answered the Lipkus, Samsa and Rimer's (2001) numeracy scale (i.e., NS), which is currently the most extensively used in research (Reyna et al., 2009). The NS is composed of the General Numeracy (GN—three items) subscale and an Expanded Numeracy (EN—eight items) subscale (for a total of 11 items). With some minor variations, the three items on the general numeracy scale correspond to those initially created by Schwartz, Woloshin, Black, and Welch (1997), with the remaining eight items (EN) added by Lipkus et al. (2001).

The participants were also asked to complete the eight-item SNS (Fagerlin, Ubel, Smith, & Zikmund-Fisher, 2007; Fagerlin, Zikmund-Fisher et al., 2007) to assess self-perception of numerical competence (Table 1). The SNS (Fagerlin, Zikmund-Fisher et al., 2007) is a self-reported measure of ability to perform mathematical tasks and preference for receiving numerical versus verbal information. The SNS consists of eight items rated on 6-point Likert-type scales, four questions asking the respondents to assess their numerical ability in different contexts (CA) and four questions asking them to state their preferences for the presentation of numerical and probabilistic information (Preference for Display of Numerical Information [PDNI]). (To avoid a plethora of abbreviations for scale names, we abbreviate the names of full scales but spell out the names of subscales.)

Cognitive Reflection Test: The Cognitive Reflection Test (CRT) is a three-item test (Table 1) that measures cognitive impulsivity or one's reliance on more automatic versus deliberative (e.g. effortful and subjectively controlled) cognitive processing (Frederick, 2005). In other words, the CRT measures the ability to suppress a spontaneous but incorrect answer.

After answering probability and memory questions, the participants responded to the Numeracy Scale (NS), to the Subjective Numeracy Scale (SNS), and finally to the Cognitive Reflection Test (CRT).

RESULTS AND DISCUSSION

NS, SNS, and CRT descriptive statistics and relationships—Means, standard deviations, and reliabilities of NS, SNS, and CRT are presented in Table 2. With regard to the three-item CRT, the percentages of the participants who gave zero, one, two, or three correct responses, respectively, were 44%, 20%, 17%, and 19%; the mean number of correct responses on the CRT was 1.10.

The objective numeracy scale (NS) demonstrated fair reliability (α=0.69; with regard to subscales of general numeracy and expanded numeracy, αs=0.60 and 0.61 respectively) and may be a more ambiguous measure in the sense that it is not grounded in empirically supported theories of numeracy or mathematical cognition. Because the questions were not theoretically motivated, they are not necessarily "pure" measures or theoretically coherent (Reyna et. al, 2009). With regard to CRT (α=0.74), although there has been theoretical progress, there has not been an extensive psychometric assessment of underlying factors.

Therefore, we performed a factor analysis encompassing the 11-item objective Numeracy Scale (NS), the eight-item Subjective Numeracy Scale (SNS) and the three-item Cognitive Reflection Test (CRT) to investigate what those tests have in common and to identify dimensions of cognitive capacity and thinking styles. We ran both oblique and orthogonal rotations. Similar results were found with both types of analyses. A non-orthogonal (oblique) solution allows the factors to be correlated; this will result in higher eigenvalues but diminished interpretability of the factors. Therefore, we present detailed results for the orthogonal solution. We used the varimax rotation that results in clearer and more comprehensible explanations of the effects we found.

The factor analysis involving NS, SNS, and CRT items resulted in six factors (Table 3). The percentage of variance accounted for by the first factor is 15.7%. Factor 2 accounted for 13.5% of the variance. Another 31.6% of the variance was accounted for by Factors 3, 4, 5, and 6. An overall examination of the factors reveals that they broke down roughly according to the preexisting scales, distinguishing objective numeracy (NS), and its subscales, from subjective numeracy (SNS), and its subscales. However, the objective numeracy subscale of General Numeracy and the CRT items loaded together on Factor 2 (but see Study 2, Table 4, in which CRT items loaded separately). Within the SNS, the Cognitive Ability and Preference for Numerical Information items generally loaded on two separate factors, Factors 1 and 6. The remaining Factors 3, 4, and 5 all derived from the objective NS (the Expanded Numeracy portion).

The working interpretations of these factors are based on which items grouped together and on a detailed examination of the errors on each item. That is, we examined commonalities among items that loaded on specific factors as well as examples of typical errors on those items and connected them to empirically supported constructs from the extant literature. The most common types of errors, the proportions of each, and the nature of the errors are presented in Tables 5 and 6. As a result of this process, we provisionally labeled the dimensions accordingly, discussed in greater detail in the succeeding paragraphs: Factor 1 is Subjective Numeracy/Abilities (CA subscale); Factor 2 is Mindless or Verbatim Matching/Monitoring; Factor 3 is Proportions; Factor 4 is Relative Magnitude (gist); Factor 5 is Multiplying; and Factor 6 is Subjective Numeracy/Preferences (PDNI subscale).

Subjective Numeracy Scale, CRT, and NS items share common variance presumably related to cognitive ability or numeracy (Table 7). However, as illustrated by the separate loading of SNS–Cognitive Ability on Factor 1, subjective self-assessments (SNS) involve judgments about one’s own cognition (metacognition), which are influenced by self-reporting biases in addition to being influenced by cognitive ability or numeracy (e.g., Dunning et al., 2004). Thus, consistent with the conclusions of Fagerlin et al. (2007), SNS is correlated with NS (0.47 in Study 1 and 0.45 in Study 2; Tables 7 and 8), but it differs in response burden (i.e., is easier to complete) and in well-known biases in self-assessment. Also, as expected, self-assessed ability (SNS–Cognitive Ability, Factor 1) does not overlap completely with self-assessed preference (SNS–PDNI, Factor 6), although, again, their raw scores are correlated (Tables 7 and 8). Cronbach’s α for these two subscales are 0.90 and 0.54, respectively. This analysis shows that the SNS (α=0.80) is indeed measuring two constructs on which each subscale loads, as intended by its creators.

For Factor 2 (Mindless or Verbatim Matching/Monitoring), all three General Numeracy items (1, 2, and 3 of NS) loaded on this dimension as did all three CRT items. In Study 2, CRT loaded separately from the General Numeracy items (Table 4). For this reason and because of the responses in Tables 5 and 6, we distinguish verbatim matching errors (characteristic of some errors on General Numeracy items, especially in Study 1) from monitoring of such errors (characteristic of some responses on CRT, especially in Study 2). We should note that these responses are less different across studies than they appear. In Study 2, for factor analyses excluding SNS items, General Numeracy items loaded 0.39, 0.42, and 0.27, respectively, on a second factor on which the CRT items also loaded (eigenvalue=1.95).

The clearest example of verbatim matching is found predominantly in Study 1 in response to NS item 1: When asked, out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)?, the most common error participants made was to answer 2, 4, or 6 (9.7% of the answers). There is little meaning attached to such unthinking responses that involve repeating information from the problem without comprehension. If you understand the meaning, you would never say 2, 4, or 6. We also describe such an answer as “mindless” (unthinking) (e.g., Reyna, 2008; Wansink, 2006). The verbatim words given in the problem are the words the participants use; they do not think deeply (mindfully) about the problem. Instead, the response is based on a mindless verbatim strategy, as found in other studies of low-level numerical and verbal reasoning (Reyna & Brainerd, 1995; Reyna et al., 2003).

The other NS items that load on this factor also elicit verbatim matching errors because they involve copying elements of the question stem to the answer blank, despite the fact that those answers do not make sense. For example, for NS item 3, when told that the chances of winning a car are 1 in 1000, and asked which percent win a car, some answer 1000 (or 1), repeating numbers in the stem but not with their correct meaning.

Similarly, when answering the CRT question, “If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets,” most people answer 100, echoing the repetition of 5, 5, 5 with 100, 100, 100. Although this strategy involves repetition of elements of the problem (100), it involves more than the simple repetition in responding 2, 4, or 6 as in NS item 1. In both problems, there is a verbatim match, but 100 is a more popular response than 2, 4, or 6, possibly increased by a kind of meaningless pattern completion and by the greater difficulty of coming up with the correct answer for the CRT problem compared with NS item 1 (see also Evans, 2003; Wason & Johnson-Laird, 1972).

Analogously, for the CRT problem in which the bat costs $1.00 more than the ball, and together they cost $1.10 (the latter is easily parsed as one dollar *and* 10 cents), many respond that the ball costs “10 cents” (or 0.10) in a kind of meaningless pattern completion. The popular response of “10 cents” in the bat-and-ball problem repeats the “missing” element of the stem. Consistent with this interpretation, people miss the “bat-and-ball” problem far more often than they miss the “banana and bagel” problem: “A banana and a bagel cost 37 cents. The banana costs 13 cents more than the bagel. How much does the bagel cost?” According to Frederick (2005), the banana and bagel problem invites computation and thus greater reflection. We would argue that this is because part of one number in the problem does not match the other number, inviting great scrutiny. If the problem specified that the banana costs 30 cents more than the bagel, subjects would be apt to respond that the bagel costs “7 cents,” an example of verbatim matching of elements of the problem.

In sum, drawing on prior research and the nature of errors, we hypothesize that Factor 2 involves using a mindless “verbatim” strategy that involves taking some element of the problem and using it in the answer without necessarily getting the gist (or meaning) of the problem. Although loading on the same factor, details of verbatim-matching strategies probably differ across questions, sometimes involving just copying information from problems and sometimes involving some additional processing (e.g., automatic computation) with the copied information. We also hypothesize that although typically reasoners first think of the wrong answer because of literal matching, some monitor their reasoning and realize that this answer is wrong and inhibit it, getting the problem right (Frederick, 2005). It is not that smart people do not think of wrong answers; they do, but then they see that they cannot be right and edit their answers (which occurred more often in Study 2).

The remaining factors were relatively robust across Studies 1 and 2. Factor 3 is called “Proportions” because it involves conversion of percentages to proportions and probabilities to proportions (NS items 8, 9, and 11, and to a lesser degree, 10). Understanding proportions, or the ratio concept, has been emphasized as a core competence in understanding probability (National Mathematics Panel, 2008; Reyna & Brainerd, 1994). Place value is the most common error participants committed when answering questions that loaded on Factor 3. Although we cannot be sure which cognitive processes participants used, their responses suggested attempts to divide one quantity by another. Wrong answers participants gave to NS item 11, for example, included 0.05 or 50 instead of 5. These responses may involve denominator neglect or a less than facile grasp of ratios because place values involve conversion of ratios (Reyna & Brainerd, 2008).

The questions that loaded on Factor 4 were NS item 4 (Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10) and item 5 (Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%). To get these problems right, respondents need only be able to put the ratios in “linear” order (e.g., from smallest to biggest) and select the “biggest.” The participants do not have to generate or compute any answers. For example, they know that, in general, 1 in 10 has to be bigger than 1 in 1000, so the other one is in the middle. There is not much variation in the answers because these problems are relatively simple and most people got them right.

Therefore, for Factor 4, the dimension is interpreted as Relative Magnitude/Gist on the basis of research showing that, beginning as young as 4years of age, relative magnitude is encoded automatically, such that differing quantities are represented along an ordinal mental number line (e.g., Bouwmeester, et al., 2007; Reyna & Brainerd, 1995; Siegler & Opfer, 2003). Gist representations of relative magnitude (ordinal gist, such as small or big) are used when answering such questions as, which is the smallest (least numerous) or biggest (most numerous) class of objects in a display (Brainerd & Gordon, 1994). Ordinal gist representations of relative magnitude are independent of verbatim representations of exact magnitudes and of exact knowledge about how to compute ratios or proportions (Reyna, et al., 2003; Reyna & Mills, 2007a).

Factor 5 is called "Multiplying" because both items that loaded on it involve a computation of doubling. NS item 6 loaded on this factor (If Person A's risk of getting a disease is 1% in 10years, and Person B's risk is double that of A's, what is B's risk?) as did NS item 7 (If Person A's chance of getting a disease is 1 in 100 in 10years, and Person B's risk is double that of A, what is B's risk?). The most common errors people made when answering these questions were dividing the number of years by two instead of doubling the risk or doubling other numbers presented in the problem. Doubling involves computational ability, or procedural knowledge of mathematics, which has been found to be distinct from conceptual knowledge, such as relative magnitude (National Mathematics, Panel, 2008; Rittle-Johnson & Siegler, 1998). In summary, factor analyses yielded six factors, two reflecting subjective numeracy, and the remaining factors provisionally interpreted as reflecting verbatim matching/monitoring, proportions (division or conversion of ratios), relative magnitude (ordinal gist), and multiplication (procedural knowledge).

Do numeracy and CRT measure the ability to make better judgments? (Study 1)—Rather than treat the foregoing scales as monolithic measures, we were interested in distinguishing how the separable factors we identified relate to reasoning performance. As discussed in the introduction, dual-process theories predict that numeracy and CRT as overall "rationality" scales should be related to biases and fallacies (e.g., Peters et al., 2006). In contrast, we relate underlying factors to biases and fallacies through regressions and show how the factors derived from the analysis described earlier predict such biases and fallacies. To do that, during the factor analysis, we saved factor scores as variables, using the regression method, and then used those as independent variables in the regression.

Our results thus far have shown that SNS loads on its own factors, which differ from those for NS (although they share variance). This result makes sense given that SNS is a self-report measure that reflects biases in self-assessment (Dunning et al., 2004). We ran regressions including SNS factors as predictors; for example, a SNS–Cognitive Ability factor predicted conjunction and disjunction biases and gist memory score, but it did not predict memory deviations. However, the inclusion of SNS predictors did not change any of the other significant predictors of reasoning biases or memory judgments. Therefore, in this section, we focus on the ability of *objective* numerical performance measures to predict biases and fallacies in probability judgment. Factor analyses with these objective measures (Table 9) yielded the same factors as before (without the SNS factors): Mindless or Verbatim Matching/Monitoring, Proportions, Relative Magnitude/Gist, and Multiplying (eigenvalues of 2.91, 1.90, 1.82, and 1.60, respectively).

The results of linear regression analyses demonstrated that Factor 1 (Verbatim Matching/ Monitoring) significantly predicted conjunction fallacies, disjunction fallacies, memory deviations, and gist memory score (Table 10). The worse the performance on the Verbatim Matching/Monitoring questions, the higher the gist memory estimate of meals (i.e., the estimated frequency of meals that were never presented, but that fit the stereotypes of the meals that were presented). The worse the performance on the Verbatim Matching/ Monitoring questions (which included the CRT items), the more conjunction fallacies, disjunction fallacies, and memory deviations people committed. The analysis of conjunction and disjunction fallacies in fuzzy-trace theory is consistent with this result, which attributes such fallacies in part to verbatim thinking (Reyna et al., 2003; Reyna & Mills, 2007a; Wolfe & Reyna, 2010). Also, conjunction and disjunction fallacies are exacerbated as a result of a failure to monitor (and adjust the probability of) intuitively compelling answers, as assessed by the CRT (see also Frederick, 2005; Reyna, 1991).

Factor 2 (Proportions) significantly predicted conjunction fallacies and memory deviations. The better the performance on the Proportions questions, the fewer conjunction fallacies and memory deviations people committed. This result is consistent with people processing conjunctive probabilities by estimating ratios or proportions. In addition, the reason why Factor 2 predicted memory deviations could be that when we asked participants how many times each meal was shown, we told them how many meals in total they had seen (20 per person). So, to estimate frequencies of presentation, they may have computed a proportion of the total number of meals that corresponded to their subjective feeling of frequency (e.g., that a meal was rarely versus frequently presented). Indeed, average estimated frequencies totaled about 20 meals, adding up estimates for tested meals that were and were not presented, as though subjects were computing a ratio using 20 as the denominator.

Factor 4 (Multiplying) also significantly predicted conjunction fallacies, which makes sense because joint probabilities can be obtained by multiplying probabilities of independent events. That is, to compute conjunctions, people might implicitly multiply the probability of the individual components of that conjunction and then adjust the result (to accommodate corrections for overlapping sets). The better people performed on the multiplying questions, the fewer conjunction fallacies they committed. Responses to Factor 3 (Relative Magnitude/ Gist) did not vary widely, and most participants reported the correct answer. This may explain why Factor 3 did not predict any bias or fallacy, but naturally, a null result is difficult to interpret. It is instructive that not all factors predicted all biases and fallacies, which allows us to begin to identify specific processes that might underlie the overall relations between numeracy and other measures of higher cognition, such as the CRT, and reasoning performance.

STUDY 2: UNITED STATES SAMPLE

A major aim of Study 2 was to determine whether the same factor structure would emerge in a different sample of respondents. To the extent that factors replicate across samples, we should place greater confidence in them as descriptions of underlying dimensions of numeracy and cognitive reflection. It is also important to determine whether any of these factors simply reflect general intelligence. If any of the factors that we have identified tap general computational ability or intelligence, they should predict reasoning errors, which we also investigate in Study 2. Furthermore, we examine whether the factors we have identified, such as monitoring, are equivalent to basic constructs such as inhibition as measured by go/ no-go tasks (response time, proportion of correct answers and proportion of false alarms were measured). Finally, we relate numeracy measures and the CRT to two scales of the REI, a well-known scale of dual processes in reasoning encompassing Need-for-Cognition (NFC, often assumed to be related to numeracy) and Faith-in-Intuition subscales (e.g., Pacini & Epstein, 1999; Reyna et al., 2009).

In this study, the participants performed another probability judgment task, a variant on the ratio-bias paradigm (Kirkpatrick & Epstein, 1992; see Reyna & Brainerd, 1994, 2008 for reviews). The ratio bias is commonly known as the tendency to judge a low probability event as more likely when presented as a large-numbered ratio, such as 10 in 100, than a smaller-numbered ratio, such as 1 in 10, even if the probabilities are the same. This bias is also known as denominator neglect, which occurs when people who understand that probability is a function of frequencies in both the numerator and the denominator still tend to pay less attention to the denominator as a default (Reyna & Brainerd, 2008).

However, people have also been shown to exhibit irrational biases in high probability contexts. For example, Kirkpatrick and Epstein (1992) showed that 63.5% of the participants preferred the small-numbered ratio in a self-perspective response when a real-life situation was simulated, despite the objective numerical equivalence of the two ratios (see also Pacini & Epstein, 1999; Reyna & Brainerd, 1994, 2008). For our purposes, we are interested in which factors predict the participants' choice of the normative response of equivalence when it is explicitly offered to them, as opposed to irrationally preferring either of the non-normative options. A ratio-bias effect has been linked to numeracy, but, again, underlying factors have not been investigated (Reyna et al., 2009).

Method

Participants—The participants were 190 Cornell University students (116 women, 74 men) ranging in age from 18 to 38years (mean=21.07, *SD*=2.78). The students were recruited in psychology courses and via campus postings. The sample was 57.9% Caucasian, 23.7% Asian-American, 5.3% Hispanic, and 13.1% African-American or mixed ethnicity. All participants gave informed consent, and the study was approved by the Institutional Review Board of Cornell University.

Procedure and material—This study was designed and run using Qualtrics.com online survey software (Qualtrics Labs Inc., Provo, UT). The respondents participated in the experiment online (and received credit toward course requirements).

In Study 2, all participants were tested on a ratio-bias problem in a high-probability winning frame:

> Two containers, labeled A and B, are filled with red and blue marbles in the following quantities.
>
> Container A contains 10 marbles, 9 red and 1 blue.
>
> Container B contains 100 marbles, 90 red and 10 blue.
>
> You must draw a marble (without looking, of course) after choosing one of the containers. If you draw a red marble, you win; otherwise you win nothing, and the game is over.

The participants were given 40 seconds to carefully read the problem. Then, they were asked on the next page to say which container gave a better chance of winning. They chose from one of three answers: (i) container A (9:10); (ii) "it would not matter to me; chances are the same"; or (iii) container B (90:100). The presentation order of the responses was randomized across subjects. The participants had no time deadline to make their choice and to reason about the problem. After performing the ratio-bias task, the participants completed the following individual difference measures.

Individual difference measures

Numeracy Scale: In this study, the participants answered the same Lipkus et al. (2001) numeracy scale (NS) that the participants answered in Study 1. In Study 2, the participants were also asked to complete the eight-item Subjective Numeracy Scale[1] (Fagerlin, Ubel et al., 2007; Fagerlin, Zikmund-Fisher et al., 2007).

Cognitive Reflection Test: In this study, we used the same CRT scale (Frederick, 2005) used in Study 1.

Cognitive capacity and rational thinking: The participants completed a 12-item short form of the Raven Advanced Progressive Matrices (APM) test (Arthur & Day, 1994) as a measure of cognitive capacity, which was designed for adults with above-average general intelligence. As a measure of rational thinking style, we used the Need-for-Cognition scale (NFC). The term “need for cognition” was defined by Cohen, Scotland, and Wolfe (1955) as “a need to understand and make reasonable the experiential world” (p. 291). Cacioppo and Petty (1982) adopted this term and proposed that need for cognition was a stable (although not invariant) individual difference in the tendency to engage in and enjoy effortful cognitive activity. More recently, Pacini and Epstein (1999) adjusted the original version of the Cacioppo and Petty scale, and this resulted in a 20-item longer form of the NFC (part of the Rational–Experiential Inventory—REI). The participants filled in this longer version of the NFC and also a 20-item Faith-in-Intuition (FI) scale (also part of the REI). The participants rated all the REI items on a 5-point scale that ranged from 1 (*definitely not true for myself*) to 5 (*definitely true for myself*). According to dual-process theories of reasoning (Evans, 2003; Kahneman & Frederick, 2002), Cognitive Ability and Need for Cognition are related to Type 2 processes (slow, controlled, limited capacity, and high effort), and some researchers have combined these two kinds of measures as an indicator of the participants’ *normative potential* (Morsanyi, Primi, Chiesi & Handley, 2009). However, cognitive capacity or intelligence is not the same thing as need for cognition, so we treated these measures separately (Reyna & Brainerd, 2007; Stanovich & West, 2008).

Go/no go: The inhibitory control ability of adults was tested with the go/no-go task (Garavan, Ross, Murphy, Roche & Stein, 2002). The participants first completed a trial version of the go/no-go task and then were tested in two blocks of stimuli. Both blocks required subjects to respond as quickly and accurately as possible by pressing the “h” key every time the “X” (go cue) appeared and not to respond to the “K” (no-go cue). Stimuli were presented in the center of the screen for 500 milliseconds. Each block contained 140 stimuli, of which 112 (80%) were go cues and 28 (20%) were no-go cues. The interstimulus interval was 500milliseconds, and the presentation order of go cues and no-go cues was pseudo randomized to discourage anticipatory responses. A fixation cross was displayed in the center of the screen during the interstimulus interval. Instructions were displayed on the computer screen at the beginning of each block, and subjects pressed the spacebar when ready to begin. Go/no-go task duration was up to 8minutes. Measures of reaction time (mean), number of correct responses, and number of false alarms were obtained for each subject.

For each scale investigating the individual differences, the presentation order of the items was randomized. In addition, the presentation order of each scale was randomized.

[1]In Study 2, we used a range of 1–5 instead of the original 1–6 (used in Study 1) to have a “midpoint” and to have a measure that is coherent and comparable with the other scales used. Correlations were apparently not affected by the change in scale and were highly similar across studies (Tables 7 and 8).

RESULTS AND DISCUSSION

NS, SNS, CRT, APM, REI, and go/no-go descriptive statistics and relationships—Means, standard deviations, and reliabilities of the NS, SNS, CRT, APM, Need-for-Cognition and Faith in Intuition (the latter two from the Rational–Experiential Inventory [REI]) are presented in Table 2. The participants performed very well on the 12-item short form of the Raven Advanced Progressive Matrices (APM) test, demonstrating a high cognitive ability. The mean proportion of correct responses was 9.31. With regard to the three-item CRT, performance was also good; the percentages of the participants who gave zero, one, two, or three correct responses, respectively, were 24%, 27%, 23%, and 26%; the mean number of correct responses to the CRT was 1.50. Interestingly, although objective performance was higher in Study 2 on NS and CRT compared with Study 1, self-assessed numeracy (SNS) was lower, further supporting the hypothesis that although objective and subjective scales share variance, they also differ in important respects. The shifting SNS scores, because they did not track objective performance, may reflect changing frames of reference, a common finding in self-assessments (Biernat, 2005).

The factor analysis involving NS, SNS, and CRT items of Study 2 produced seven factors (Table 4). The percentage of variance accounted for by the first factor is 15.6%. Factor 2 accounts for 9.5% of the variance. Another 35.9% of the variance is accounted by Factors 3, 4, 5, 6, and 7. The dimensions that emerged in Study 2 were similar to those of Study 1. As before, Factor 1 reflected SNS items (mainly Cognitive Ability, but also two PDNI items had loadings of 0.61 and 0.48 on this factor). The remaining two SNS–PDNI items loaded on Factor 4.

As also shown in Table 4, Factor 2 reflected Proportions, with NS items 8, 9, and 10 loading on this factor, along with NS item 1, probably because the latter item was now responded to by calculating a proportion (i.e., 50% of 1000 or 500 rolls) rather than by verbatim matching (e.g., 2, 4, or 6; see responses described in Tables 5 and 6). The CRT items loaded on a third factor (Monitoring), separate from General Numeracy items, although as noted earlier, all three CRT items loaded somewhat on Factor 3 (along with the CRT items) when SNS items were excluded, resembling results in Study 1. As in Study 1, there was one factor interpretable as Multiplying (NS items 4 and 5, Factor 6) and another factor interpretable as Relative Magnitude/Gist (NS items 6 and 7, Factor 7). The main difference between this study and Study 1 was that CRT items loaded alone; NS items (3 and 11) loaded together on a separate factor, which might reflect verbatim matching but was difficult to interpret.

In Study 2, the reason why NS items loaded on a different factor than Monitoring (CRT items), and some of the NS items changed factors, is probably that the participants in Study 2 not only scored higher on the NS (with almost no variance on some items) but also committed different kinds of errors, compared with the participants in Study 1. The participants in Study 2 made fewer literal or verbatim matching errors on NS items than the participants did in Study 1 (Tables 5 and 6). In the dice problem (NS item 1), for example, the participants in Study 2 committed matching errors the same percentage of times that they committed place-value errors, whereas in Study 1, matching errors were eight times more frequent than place-value errors. That change presumably moved NS item 1 from the Verbatim Matching factor in Study 1 to the Proportions factor in Study 2.

Although the loadings of SNS subscale items seem to be somewhat different in Study 2 compared with Study 1 at first glance, closer inspection of the loadings reveals that they are similar across studies. For example, SNS–PDNI items 1, 2, and 4 loaded 0.47, 0.28, and 0.48, respectively, on Factor 1 in Study 1, and these same items loaded 0.48, 0.36, and 0.61 on Factor 1 in Study 2 (Tables 3 and 4). SNS–PDNI items 2 and 3 loaded together on Factor 6 in Study 1 and on Factor 4 in Study 2 (Tables 3 and 4). In Study 1, the NS preceded the

SNS, which one might argue could contribute to fluency effects. However, in Study 2, the order of scales was randomized, with similar findings as in Study 1, indicating that processing fluency did not appreciably affect the factor solution obtained in Study 1. Also, SNS scores and Need-for-Cognition scores correlated reasonably well (r=0.54), which makes sense because both are subjective self-reports, and both tests measure how much people like to deal with challenging tasks.

Study 2 allows us to go beyond Study 1 in the important respect of determining whether any of our measures, or factors derived from those measures, are equivalent to general intelligence. Study 2 also allows us to determine whether validated dual-process measures, such as those on the REI, relate to numeracy as hypothesized in prior work. As shown in Table 8, APM correlated positively with NS, CRT, NFC, the SNS–Cognitive Ability factor, and the Relative Magnitude/Gist factor (NS items 6 and 7). APM correlated negatively with Faith in Intuition (FI) and with false alarms in the go/no-go task. However, none of these correlations are sufficiently high to justify the conclusion that the measures are redundant with intelligence. Adding the APM score along with Need for Cognition, Faith in Intuition, and the go/no-go measures in a factor analysis with NS, SNS, and CRT items produced separate factors for Faith in Intuition and for the go/no-go measures (which loaded together), whereas the other factors remained the same (Table 11). In this analysis, the APM score loaded on the same factor (loading of 0.47) as the Relative Magnitude/Gist items (NS items 6 and 7), as fuzzy-trace theory would expect because gist representations are associated with higher levels of cognition (Reyna & Brainerd, 1995; Reyna & Lloyd, 2006; Reyna et al., 2003). Similar factors are extracted for NS and CRT items when SNS items are removed (Table 12).

Do numeracy and CRT measure the ability to make better judgments? (Study 2)—Logistic regression was used to investigate how the factors derived from the factor analyses predicted the responses given to the ratio-bias task (Table 13). The SNS factors did not significantly predict ratio bias, and the inclusion of the SNS factor scores as predictors did not change which remaining predictors were significant, as in Study 1. Therefore, Table 13 presents the results of the regression analysis for the factors derived from the objective measures. The right side of Table 13 shows the odds ratios (correct/wrong responses), which measures effect size, and the antilog (i.e., exponentiated values) of the model coefficients. Factor 1, Proportions, and Factor 5, Relative Magnitude/Gist predicted normative responding on the ratio-bias task; scoring higher on proportions was associated with correct ratio responses, as might be expected, as was scoring higher on judging relative magnitudes of ratios, which also makes sense. Monitoring, Multiplying, and Verbatim Matching were not associated with ratio-bias performance.

GENERAL RESULTS AND DISCUSSION

Do subjective and objective numeracy measure the same construct?

Subjective numeracy measures were developed to assess how confident and comfortable people feel about their ability to understand and apply numbers without actually having to perform any numerical operations. Subjective measures increase the convenience and acceptability of measuring numeracy for respondents, relative to objective measures that usually require strenuous effort and are potentially aversive. The goal was to create a measure to allow subjective numeracy to substitute for objective numeracy when the latter is not practical, but the extent of overlap in the constructs has not been extensively studied (Reyna et al., 2009).

We investigated the question of whether objective and subjective numeracy scales measure the same construct, and the answer is that, although they share variance, they also differ in important ways. The correlations between NS and SNS observed in both studies, despite differences in the samples, were virtually identical and about 0.20 lower than that reported by Fagerlin et al. (2007; 0.45 and 0.47 compared with 0.68). Not only did the scales not correlate very highly with one another, but their test items did not load on the same factors. The relatively low correlations between objective performance and subjective self-assessment, the higher performance in the sample which rated itself as lower subjectively, and the failure to load on the same factors all suggest that people are relatively poor judges of their ability to understand and use numbers and that SNS is not entirely a substitute for NS. In clinical and practical settings, however, the Subjective Numeracy Scale has positive features. Unfortunately, subjective and objective numeracy did not correlate well enough with each other to be interchangeable (especially SNS and Expanded Numeracy items; Tables 7 and 8), and their items loaded on separate factors in two independent samples. In fact, in Study 1, CRT correlated slightly higher with objective numeracy than subjective numeracy did.

The smaller correlations between objective and subjective numeracy observed here could also be explained by the fact that Fagerlin et al. (2007) had a community sample with a wider range of numeracy scores than those of college students (although Study 1 had lower performance than Study 2 and was a large sample). Low correlations in the present, more homogeneous college student samples are consistent with the finding of Fagerlin et al. (2007) of higher correlations in a more heterogeneous population of hospital visitors. Nevertheless, for the populations studied, the two measures are distinguishable. These results open the door to further research identifying how self-report biases, including shifting frames of reference, influence subjective numeracy scores, and how such biases can be reduced.

Is the Cognitive Reflection Test just another objective numeracy test?

One might think that CRT is just another numeracy scale because the questions are numerical. Campitelli and Labollita (2010) pointed out that CRT might be measuring numeracy, but noted that it correlates with performance in a task without mathematics. They ultimately concluded that CRT taps a broader concept of actively open-minded thinking. Actively open-minded thinking allows people to generate different answers to a question. Thus, in numeracy tasks, actively open-minded people who also have computational ability can generate many candidate answers to a problem. In this way, actively open-minded thinking would be expected to be related to computational performance (e.g., in mental arithmetic; Reyna & Brainerd, 1995). The ease and automaticity of computation for some individuals can also lure them into mindless computation (e.g., computing proportions when simpler arithmetical operations are correct) because they are good and fast at it (Reyna et al., 2009).

However, to answer questions correctly, it is not enough to have actively open-minded thinking or computational ability. One must edit out the wrong answers from the many that are generated. That is, one might realize that an answer is wrong (reflection), inhibit, and edit it (Frederick, 2005). For example, most people generate the immediate answer (10 cents) with the bat-and-ball problem, and then they have to check their answer to get the right answer. People usually think of the wrong answer first on the basis of literal matching, and then they have to inhibit this wrong answer; they must withhold the “mindless verbatim” answer. It makes sense that smart people would inhibit wrong answers and that this would be related to intelligence. Consistent with this interpretation, in Study 2, the APM score correlated with CRT performance, as reported in previous work. However, CRT and inhibition tasks (go/no go) did not correlate with each other, which suggests that monitoring (a metacognitive judgment, like subjective numeracy) is distinct from inhibition, at least from behavioral inhibition (Reyna & Mills, 2007b). The finding that CRT correlates with

Need for Cognition also supports this interpretation that CRT captures monitoring, and editing, responses (e.g., actively engaging in computation).

Cognitive Reflection Test answers obtained in our studies were very similar to the answers obtained by Frederick (2005). Most people get the questions wrong, and they give the “mindless” and sometimes “verbatim” answers noted earlier. There is one subtle distinction that we want to make clear, however. Frederick (2005) implies that those common wrong answers that people give are intuitive. However, there are “dumb” intuitive and “smart” intuitive answers. Dumb intuition is just looking at some information in the problem and matching it verbatim; it is literal. In this context, Frederick’s explanation makes sense. When most researchers use the term “intuitive,” that is the sense they mean.

In fuzzy-trace theory, there are multiple kinds of intuitive answers (Reyna, 2004). According to this theory of intuition, the kind of common wrong answers that people give when answering CRT questions would not be called “intuitive” but rather would be called verbatim, the kind of answer that people give when they lack comprehension or do not think hard enough to answer correctly (e.g., Reyna et al., 2003). One of the main tenets of the theory is that advanced cognition is typically gist-based intuition. A fuzzy-processing preference (i.e., a preference for gist-based intuition) increases with age from childhood to adulthood and with increasing expertise in adulthood (Reyna, 2008; Reyna et al., in press; Reyna & Lloyd, 2006). A gist-based intuitive answer, then, would be one based in gist memories, a kind of information that people retain after understanding something and giving meaning to it. Consistent with this interpretation, in our study, the CRT is not correlated with the Faith in Intuition score (Table 8) that is posited to measure primitive “intuitive answers.”

In sum, to determine whether CRT is just another numeracy scale, we computed bivariate correlations between the scales and then performed factor analyses on items. The two scales were related, especially the general numeracy and CRT items but not reliably across studies. The bivariate correlations between them were not high, indicating that they were not equivalent, but the restricted range of scores for the CRT would attenuate such a correlation. Thus, the answer to the question of whether CRT is just another numeracy test would seem to be a qualified no. Given the correlations between CRT and intelligence, and between intelligence and numeracy, it would be important to control for intelligence in future research linking these concepts.

SUMMARY AND CONCLUDING DISCUSSION

Despite the fact that our two samples were obtained from different countries, the results were remarkably similar: Items on the objective numeracy, subjective numeracy, and the CRT could be grouped into interpretable dimensions on the basis of factor analyses. These factors separated the ability to extract the relative gist of quantities from verbatim matching of elements of problems (and failure to monitor and censor these verbatim answers), and each of these was separated from computational skills, such as computing proportions (ratios) and multiplying. These factors successfully predicted memory performance, conjunction and disjunction fallacies in reasoning about probabilities, as well as ratio bias in probability judgment. The results were generally consistent with dual-process approaches to reasoning and probability judgment, in particular, the distinction between verbatim-based and gist-based processes (e.g., Epstein, 1994; Evans, 2007; Kahneman & Frederick, 2002; Klaczynski & Cottrell, 2004; Kühberger & Tanner, 2010; Reyna, 2004).

The observed factors, and the associated pattern of predictions, begin to illuminate the cognitive dimensions underlying numeracy tests and the CRT. Specifically, the results support the conclusions that the CRT is not just another numeracy scale; that objective and subjective numeracy scales overlap but differ in important ways; and that multiple factors captured in numeracy scales predict biases and fallacies in probability judgment. These results advance our understanding of the cognitive mechanisms captured in assessments of numeracy and how they relate to cognitive theory.

Acknowledgments

CAPES (Brazil) supported part of this research.

References

Arthur W Jr, Day DV. Development of a short form for the Raven advanced progressive matrices test. Educational and Psychological Measurement. 1994; 54:394–403.

Bechara A, Damasio H, Tranel D, Damasio AR. Deciding advantageously before knowing the advantageous strategy. Science. 1997; 275:1293–1294. [PubMed: 9036851]

Biernat, M. Standards and expectancies: Contrast and assimilation in judgments. New York: Psychology Press/Taylor and Francis; 2005.

Bouwmeester S, Vermunt JK, Sijtsma K. Development and individual differences in transitive reasoning: A fuzzy trace theory approach. Developmental Review. 2007; 27(1):41–74.

Brainerd CJ, Gordon LL. Development of verbatim and gist memory for numbers. Developmental Psychology. 1994; 30(2):163–177.

Brainerd CJ, Reyna VF. Explaining memory-free reasoning. Psychological Science. 1992; 3:332–339.

Cacioppo JT, Petty RE. The need for cognition. Journal of Personality and Social Psychology. 1982; 42:116–131.

Campitelli G, Labollita M. Correlations of cognitive reflection with judgments and choices. Judgment and Decision Making. 2010; 3:182–191.

Cohen AR, Scotland E, Wolfe DM. An experimental investigation of need for cognition. Journal of Abnormal and Social Psychology. 1955; 51:291–294.

Cokely ET, Kelley CM. Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. Judgment and Decision Making. 2009; 4:20–33.

Dunning D, Heath C, Suls J. Flawed self-assessment: Implications for health, education, and the workplace. Psychological Science in the Public Interest. 2004; 5:69–106.

Epstein S. Integration of the cognitive and the psychodynamic unconscious. American Psychologist. 1994; 49:709–724. [PubMed: 8092614]

Epstein S, Pacini R, Denes-Raj V, Heier H. Individual differences in intuitive–experiential and analytical–rational thinking styles. Journal of Personality and Social Psychology. 1996; 71:390–405. [PubMed: 8765488]

Evans JSBT. In two minds: Dual-process accounts of reasoning. Trends in Cognitive Sciences. 2003; 7:454–459. [PubMed: 14550493]

Evans, JSBT. Thinking: Dual processes in reasoning and judgment. Hove: Psychology Press; 2007.

Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods. 1999; 4:272–299.

Fagerlin A, Ubel PA, Smith DM, Zikmund-Fisher BJ. Making numbers matter: Present and future research in risk communication. American Journal of Health Behavior. 2007; 31:S47–S56. [PubMed: 17931136]

Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: Development of the subjective numeracy scale. Medical Decision Making. 2007; 27:672–680. [PubMed: 17641137]

Frederick S. Cognitive reflection and decision making. Journal of Economic Perspectives. 2005; 19:25–42.

Garavan H, Ross TJ, Murphy K, Roche RAP, Stein EA. Dissociable executive functions in the dynamic control of behavior: Inhibition, error detection, and correction. Neuro-Image. 2002; 17:1820–1829. [PubMed: 12498755]

Kahneman D. A perspective on judgment and choice: Mapping bounded rationality. American Psychologist. 2003; 58:697–720. [PubMed: 14584987]

Kahneman, D.; Frederick, S. Representativeness revisited: Attribute substitution in intuitive judgment. In: Gilovich, T.; Griffin, D.; Kahneman, D., editors. Heuristics and biases: The psychology of intuitive judgment. New York: Cambridge University Press; 2002. p. 49-81.

Kirkpatrick LA, Epstein S. Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. Journal of Personality and Social Psychology. 1992; 63:534–544. [PubMed: 1447684]

Klaczynski PA, Cottrell JE. A dual-process approach to cognitive development: The case of children's understanding of sunk cost decisions. Thinking & Reasoning. 2004; 10:147–174.

Kühberger A, Tanner C. Risky choice framing: Task versions and a comparison of prospect-theory and fuzzy-trace theory. Journal of Behavioral Decision Making. 2010; 23:314–329.

Lipkus IM, Peters E. Understanding the role of numeracy in health: Proposed theoretical framework and practical insights. Health Education & Behavior. 2009; 36:1065–1081. [PubMed: 19834054]

Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Medical Decision Making. 2001; 21:37–44. [PubMed: 11206945]

Milkman KL, Chugh D, Bazerman MH. How can decision making be improved? Perspectives on Psychological Science. 2009; 4:379–383.

Morsanyi K, Primi C, Chiesi F, Handley S. The effects and side-effects of statistics education: Psychology students' (mis-)conceptions of probability. Contemporary Educational Psychology. 2009; 34:210–220.

National Mathematics Advisory Panel. Foundations for success: The final report of the National Mathematics Advisory Panel. U.S. Department of Education; Washington: 2008.

Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: Theory and practice. Annals of Behavioral Medicine. 2008; 35:261–274. [PubMed: 18677452]

Oechssler J, Roider A, Schmitz PW. Cognitive abilities and behavioral biases. Journal of Economic Behavior and Organization. 2009; 29:147–152.

Pacini R, Epstein S. The relation of rational and experiential information processing style, basic beliefs, and the ratio-bias phenomenon. Journal of Personality and Social Psychology. 1999; 76:972–987. [PubMed: 10402681]

Pedhazur, EJ.; Schmelkin, LP. Measurement, design, and analysis: An integrated approach. Hillsdale: Lawrence Erlbaum Associates; 1991.

Peters E, Västfjäll D, Slovic P, Mertz C, Mazzocco K, Dickert S. Numeracy and decision making. Psychological Science. 2006; 17:407–413. [PubMed: 16683928]

Reyna VF. Class inclusion, the conjunction fallacy, and other cognitive illusions. Developmental Review. 1991; 11:317–36.

Reyna VF. How people make decisions that involve risk: A dual-processes approach. Current Directions in Psychological Science. 2004; 13:60–66.

Reyna VF. A theoryof medicaldecision making andhealth: Fuzzy-trace theory. Medical Decision Making. 2008; 28:850–865. [PubMed: 19015287]

Reyna, VF.; Brainerd, CJ. The origins of probability judgment: A review of data and theories. In: Wright, G.; Ayton, P., editors. Subjective probability. New York: Wiley; 1994.

Reyna VF, Brainerd CJ. Fuzzy-trace theory: An interim synthesis. Learning and Individual Differences. 1995; 7:1–75.

Reyna VF, Brainerd CJ. The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. Learning and Individual Differences. 2007; 17:147–159.

Reyna VF, Brainerd CJ. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. Learning and Individual Differences. 2008; 18:89–107.

Reyna VF, Farley F. Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. Psychological Science in the Public Interest. 2006; 7:1–44.

Reyna VF, Kiernan B. The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instruction, and retention interval. Developmental Psychology. 1994; 30:178–191.

Reyna VF, Lloyd FJ. Physician decision-making and cardiac risk: Effects of knowledge, risk perception, risk tolerance, and fuzzy processing. Journal of Experimental Psychology. Applied. 2006; 12:179–195. [PubMed: 16953744]

Reyna VF, Mills BA. Converging evidence supports fuzzy-trace theory's nested sets hypothesis (but not the frequency hypothesis). The Behavioral and Brain Sciences. 2007a; 30:278–280.

Reyna, VF.; Mills, BA. Interference processes in fuzzy-trace theory: Aging, Alzheimer's disease, and development. In: MacLeod, C.; Gorfein, D., editors. Inhibition in cognition. Washington: APA Press; 2007b. p. 185-210.

Reyna VF, Estrada SM, DeMarinis JA, Myers RM, Stanisz JM, Mills BA. Neurobiological and memory models of risky decision making in adolescents versus young adults. Journal of Experimental Psychology. Learning, Memory, and Cognition. (In press).

Reyna, VF.; Lloyd, FJ.; Brainerd, CJ. Memory, development, and rationality: An integrative theory of judgment and decision making. In: Schneider, S.; Shanteau, J., editors. Emerging perspectives on judgment and decision research. New York: Cambridge University Press; 2003. p. 201-245.

Reyna VF, Nelson WL, Han PK, Dieckmann NF. How numeracy influences risk comprehension and medical decision making. Psychological Bulletin. 2009; 135:943–973. [PubMed: 19883143]

Rittle-Johnson, B.; Siegler, RS. The relation between conceptual and procedural knowledge in learning mathematics: A review of the literature. In: Donlan, C., editor. The development of mathematical skills. East Sussex: Psychology Press; 1998. p. 75-110.

Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Annals of Internal Medicine. 1997; 127(11):966–972. [PubMed: 9412301]

Siegler RS, Opfer J. The development of numerical estimation: Evidence for multiple representations of numerical quantity. Psychological Science. 2003; 14(3):237–243. [PubMed: 12741747]

Stanovich KE, West RF. Individual differences in reasoning: Implications for the rationality debate? The Behavioral and Brain Sciences. 2000; 23:645–726. [PubMed: 11301544]

Stanovich KE, West RF. On the relative independence of thinking biases and cognitive ability. Journal of Personality and Social Psychology. 2008; 94:672–695. [PubMed: 18361678]

Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science. 1974; 185:1124–1131. [PubMed: 17835457]

Tversky A, Kahneman D. Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review. 1983; 90:293–315.

Wansink, B. Mindless eating: Why we eat more than we think. New York: Bantam Dell; 2006.

Wason, P.; Johnson-Laird, P. Psychology of reasoning: Structure and content. Cambridge: Harvard University Press; 1972.

Weber EU, Johnson E. Mindful judgment and decision making. Annual Review of Psychology. 2009; 60:53–85.

Wolfe CR, Reyna VF. Semantic coherence and fallacies in estimating joint probabilities. Journal of Behavioral Decision Making. 2010; 23:203–223.

Biographies

Jordana M. Liberali is a Ph.D. student of marketing at the Erasmus School of Economics of the Erasmus University Rotterdam and a doctoral candidate in Psychology at PUCRS, Brazil. In 2010 she was a visiting student at the Laboratory of Rational Decision Making at Cornell University. Her research interests include the influence of cognitive factors on judgment and decision making, individual differences in decision-making competence and memory.

Valerie F. Reyna is Professor and Co-Director of the Center for Behavioral Economics and Decision Research at Cornell University. She is a developer of fuzzy-trace theory, a model of memory and decision-making, applied in law, medicine, and public health. Her recent work has focused on numeracy, medical decision-making, risk perception and risk-taking, neurobiological models of development, and neurocognitive impairment and genetics. Past President of the Society for Judgment and Decision Making, she serves on advisory boards for the National Academy of Sciences and the Food and Drug Administration.

Sarah Furlan studied Psychology with specialization in Experimental and Methodological Psychology at the University of Padua (Italy) before receiving her Ph.D. there in Developmental Psychology in 2011. In 2010 she was a visiting scholar at the Laboratory of Rational Decision Making at Cornell University where she focused on aspects related to numeracy and decision-making based on probability information. Dr. Furlan is currently a post-doctoral Fellow in the Department of Developmental Psychology and Socialization at the University of Padua. Her research interests include probability judgments and statistical reasoning in children, adolescents, and adults; development of everyday-life beliefs connected to luck and chance and advances in data analysis in the behavioral sciences.

Lilian M. Stein Ph.D. is Professor of Human Cognition at the Psychology Postgraduate Department of the Pontifical Catholic University of Rio Grande do Sul (PUCRS), Brazil. Accredited investigator (level 1) by the Brazilian National Council for Scientific and Technological Development (CNPq). Her research interests focuses on the effect of emotion on memory and judgment and decision making.

Seth T. Pardo Ph.D. earned his Doctorate in Developmental Psychology from the Department of Human Development, Cornell University, Ithaca, NY. His research focuses on identity development and decision making.

APPENDIX: FACTOR ANALYSIS ROTATION DISCUSSION

With regard to factor analysis, Fabrigar, Wegener, MacCallum, and Strahan (1999) recommend an oblique rotation rather than an orthogonal solution. They note that dimensions of interest to psychologists are not often the dimensions we would expect to be orthogonal. If the latent variables are, in fact, correlated, then an oblique rotation will produce a better estimate of the true factors than will an orthogonal rotation. If the oblique rotation indicates that the factors have close to zero correlations, however, the analyst can go ahead and conduct an orthogonal rotation (which should then give about the same solution as the oblique rotation). Pedhazur and Schmelkin (1991) agree that if the oblique rotation demonstrates a negligible correlation between the extracted factors, then it is reasonable to use the orthogonally rotated solution.

Table 1

Items on the Objective Numeracy Scale, Subjective Numeracy Scale, and the Cognitive Reflection Test

Lipkus et al., (2001) Objective Numeracy Scale (NS)	
1	Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)?
2	In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS?
3	In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?
4	Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10
5	Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%
6	If Person A's risk of getting a disease is 1% in 10years, and Person B's risk is double that of A's, what is B's risk?
7	If Person A's chance of getting a disease is 1 in 100 in 10years, and Person B's risk is double that of A, what is B's risk?
8	If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?
9	If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?
10	If the chance of getting a disease is 20 out of 100, this would be the same as having a ____ % chance of getting the disease.
11	The chance of getting a viral infection is 0.0005. Out of 10 000 people, about how many of them are expected to get infected?
Fagerlin et al., (2007) **Subjective Numeracy Scale (SNS)** *Cognitive abilities*	
1	How good are you at working with fractions? (1=not at all good, 6=extremely good)
2	How good are you at working with percentages? (1=not at all good, 6=extremely good)
3	How good are you at calculating a 15% tip? (1=not at all good, 6=extremely good)
4	How good are you at figuring out how much a shirt will cost if it is 25% off? (1=not at all good, 6=extremely good)
Preference for Display of Numerical Information	
5	When reading the newspaper, how helpful do you find tables and graphs that are parts of a story? (1=not at all, 6=extremely)
6	When people tell you the chance of something happening, do you prefer that they use words ("it rarely happens") or numbers ("there's a 1% chance")? (1=always prefer words, 6=always prefer numbers)
7	When you hear a weather forecast, do you prefer predictions using percentages (e.g., "there will be a 20% chance of rain today") or predictions using only words (e.g., "there is a small chance of rain today")? (1=always prefer percentages, 6=always prefer words; reverse coded)
8	How often do you find numerical information to be useful? (1=never, 6=very often)
Frederick's (2005) **Cognitive Reflection Test (CRT)**	
1	A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? ____ cents
2	If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes
3	In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days

Table 2

Means, standard deviations, and reliabilities for the NS, SNS, CRT, REI, APM, go/no-go, and subscales

	Min	Max	*M*	*SD*	*Alpha*
Study 1					
Objective Numeracy Scale					
Objective Numeracy (total)	1	11	8.90	1.99	0.69
General Numeracy	0	3	2.15	0.98	0.60
Expanded Numeracy	1	8	6.73	1.39	0.61
Subjective Numeracy Scale					
Subjective Numeracy (total)	2.00	6	4.24	0.85	0.80
Cognitive Abilities	1.25	6	4.23	1.08	0.90
Preference for Display of Numerical Information	1.75	6	4.24	0.92	0.54
Cognitive Reflection Test					
CRT score	0	3	1.10	1.17	0.74
Study 2					
Objective Numeracy Scale					
Objective Numeracy (total)	2	11	9.95	1.35	0.59
General Numeracy	0	3	2.43	0.79	0.44
Expanded Numeracy	2	8	7.52	0.83	0.46
Subjective Numeracy Scale					
Subjective Numeracy (total)	1.63	5	3.72	0.67	0.79
Cognitive Abilities	1	5	3.71	0.83	0.82
Preference for Display of Numerical Information	1.75	5	3.73	0.74	0.61
Cognitive Reflection Test					
CRT score	0	3	1.50	1.12	0.64
Rational–Experiential Inventory					
Need-for-Cognition Rationality (total)	41	100	73.08	11.04	0.89
Rational ability	20	50	36.00	6.07	0.84
Rational engagement	21	50	37.08	6.23	0.84
Faith-in-Intuition Experientiality (total)	41	92	66.90	10.57	0.89
Experiential ability	19	47	33.54	5.54	0.81
Experiential engagement	17	48	33.36	6.11	0.84

	Min	Max	*M*	*SD*	*Alpha*
Advanced Progressive Matrices (APM)					
APM score	1	12	9.31	2.34	0.74
Go/no go					
Reaction time	167.42	499.27	310.69	51	–
Proportion of correct	0.20	0.98	0.85	0.15	–
Proportion of false alarms	0	0.24	0.07	0.04	–

Note: The maximum for Faith-in-Intuition scale is 92, and the maximum for components (Experiential Ability+Experiential Engagement) adds to 95 (47+48). This is because in this case, the person who adds to 47 in the ability is not the same person who adds to 48 in the engagement, so the total is not the sum of the subscales. *SD*, standard deviation.

Table 3

Study 1. Factor loadings for factor analysis with varimax rotation for NS, SNS, and CRT

	Components					
Items for each scale	**1**	**2**	**3**	**4**	**5**	**6**
CRT 1—A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?	0.16	**0.71**	−0.08	0.10	−0.03	0.05
CRT 2—If it takes 5 machines 5minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?	0.25	**0.70**	0.08	−0.09	0.02	−0.08
CRT 3—In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?	0.26	**0.71**	0.18	−0.06	0.09	0.06
NS 1—Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)?	0.17	**0.51**	0.08	0.25	0.15	0.22
NS 2—In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS?	−0.03	**0.54**	0.12	0.14	0.22	0.31
NS 3—In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?	0.14	**0.61**	−0.01	0.01	0.15	0.18
NS 4—Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10	0.07	0.09	0.10	**0.89**	0.03	0.04
NS 5—Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%	0.14	−0.01	0.14	**0.85**	−0.01	0.03
NS 6—If Person A’s risk of getting a disease is 1% in 10years, and Person B’s risk is double that of A’s, what is B’s risk?	0.13	0.12	0.08	0.01	**0.82**	−0.07
NS 7—If Person A’s chance of getting a disease is 1 in 100 in 10years, and Person B’s risk is double that of A, what is B’s risk?	0.15	0.17	0.06	0.03	**0.83**	0.04
NS 8—If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?	0.04	0.05	**0.79**	0.26	0.16	0.03
NS 9—If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?	−0.02	−0.05	**0.84**	0.06	0.07	0.21
NS 10—If the chance of getting a disease is 20 out of 100, this would be the same as having a ____% chance of getting the disease.	0.17	0.32	0.39	0.34	0.15	−0.05
NS 11—The chance of getting a viral infection is 0.0005. Out of 10000 people, about how many of them are expected to get infected?	0.12	0.29	**0.50**	−0.08	−0.18	− 0.29
SNS–CA 1—How good are you at working with fractions?	**0.71**	0.32	0.07	0.08	0.19	0.14
SNS–CA 2—How good are you at working with percentages?	**0.80**	0.30	0.07	0.08	0.11	0.09
SNS–CA 3—How good are you at calculating a 15% tip?	**0.86**	0.17	−0.03	0.09	0.12	0.10
SNS–CA 4—How good are you at figuring out how much a shirt will cost if it is 25% off?	**0.85**	0.12	0.00	0.12	0.06	0.01
SNS–PDNI 1—When reading the newspaper, how helpful do you find tables and graphs that are parts of a story?	0.47	0.14	0.12	0.04	−0.02	0.38
SNS–PDNI 2—When people tell you the chance of something happening, do you prefer that they use words (“it rarely happens”) or numbers (“there’s a 1% chance”)?	0.28	0.22	0.10	0.02	0.09	**0.59**
SNS–PDNI 3—When you hear a weather forecast, do you prefer predictions using percentages (e.g., “there will be a 20% chance of rain today”) or predictions using only words (e.g., “there is a small chance of rain today”)?	−0.01	0.07	−0.07	0.04	−0.06	**0.63**
SNS–PDNI 4—How often do you find numerical information to be useful?	0.48	0.11	0.16	−0.05	−0.08	**0.58**
Eigenvalues	3.45	2.98	1.90	1.85	1.62	1.59

Note: Factor loadings >0.50 are in boldface. CRT, Cognitive Reflection Test; NS, Objective Numeracy Scale; SNS, Subjective Numeracy Scale; CA, Subjective Numeracy's subscale Cognitive Ability; PDNI, Subjective Numeracy's subscale Preference for Display Numerical Information; Factor 1 is Subjective Numeracy/Abilities (Cognitive Abilities Subscale); Factor 2 is Literal or Verbatim Thinking/Monitoring; Factor 3 is Proportions; Factor 4 is Relative Magnitude/Gist; Factor 5 is Multiplying, and Factor 6 is Subjective Numeracy/Preferences (Preference for Display of Numerical Information subscale).

Table 4

Study 2. Factor loadings for factor analysis with varimax rotation for the NS, SNS, and CRT

	Components						
Item for each scale	**1**	**2**	**3**	**4**	**5**	**6**	**7**
CRT 1—A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?	0.09	−0.04	**0.77**	0.07	−0.02	0.04	0.01
CRT 2—If it takes 5 machines 5minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?	0.16	−0.02	**0.75**	−0.03	0.18	0.00	0.12
CRT 3—In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?	0.29	0.16	**0.56**	0.20	0.29	0.01	0.11
NS 1—Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)?	0.17	**0.54**	0.21	0.48	−0.17	− 0.06	0.02
NS 2—In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS?	0.17	0.35	0.21	0.44	0.31	−0.01	−0.29
NS 3—In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?	0.27	0.09	0.16	0.08	**0.61**	−0.02	0.06
NS 4—Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10	−0.03	0.44	0.15	0.16	−0.05	0.00	**0.61**
NS 5—Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%	0.10	−0.08	0.09	0.01	−0.02	−0.07	**0.77**
NS 6—If Person A's risk of getting a disease is 1% in 10years, and Person B's risk is double that of A's, what is B's risk?	0.06	−0.04	−.11	0.05	0.07	**0.88**	0.02
NS 7—If 'Person's chance of getting a disease is 1 in 100 in 10years, and Person B A 's risk is double that of A, what is B's risk?	0.08	0.20	0.18	0.11	0.00	**0.79**	−0.10
NS 8—If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?	0.00	**0.77**	−0.07	−0.12	0.22	− 0.04	0.13
NS 9—If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?	0.29	**0.53**	−0.03	−0.12	0.34	0.05	0.07
NS 10—If the chance of getting a disease is 20 out of 100, this would be the same as having a ____% chance of getting the disease.	0.05	**0.63**	0.01	−0.03	− 0.02	0.15	−0.10
NS 11—The chance of getting a viral infection is 0.0005. Out of 10000 people, about how many of them are expected to get infected?	−0.04	0.10	0.10	0.05	**0.70**	0.06	−0.11
SNS–CA 1—How good are you at working with fractions?	**0.70**	0.09	0.19	−0.02	0.30	0.02	0.07
SNS–CA 2—How good are you at working with percentages?	**0.87**	0.03	0.01	0.09	0.12	−0.05	0.14
SNS–CA 3—How good are you at calculating a 15% tip?	**0.73**	0.19	0.22	0.08	−0.03	−0.05	−0.25
SNS–CA 4—How good are you at figuring out how much a shirt will cost if it is 25% off?	**0.76**	0.18	0.21	0.12	−0.10	0.14	−0.14
SNS–PDNI 1—When reading the newspaper, how helpful do you find tables and graphs that are parts of a story?	0.48	−0.21	−0.03	0.24	0.34	0.18	0.21
SNS–PDNI 2—When people tell you the chance of something happening, do you prefer that they use words ("it rarely happens") or numbers ("there's a 1% chance")?	0.36	−0.06	0.13	**0.60**	−0.07	0.14	0.05
SNS–PDNI 3—When you hear a weather forecast, do you prefer predictions using percentages (e.g., "there will be a 20% chance of rain today") or predictions using only words (e.g., "there is a small chance of rain today")?	0.03	−0.14	−0.04	**0.75**	0.18	0.09	0.10
SNS–PDNI 4—How often do you find numerical information to be useful?	**0.61**	0.00	0.09	0.29	0.08	0.15	0.26
Eigenvalues	3.43	2.10	1.82	1.63	1.55	1.54	1.36

Note: Factor loadings >0.50 are in boldface. CRT, Cognitive Reflection Task; NS, Objective Numeracy Scale; SNS, Subjective Numeracy Scale; CA, Subjective Numeracy's subscale Cognitive Ability; PDNI, Subjective Numeracy's subscale Preference for Display Numerical Information; Factor 1 is Subjective Numeracy Cognitive Abilities; Factor 2 is Proportions; Factor 3 is Cognitive Reflection Test; Factor 4 is SNS–PDNI; Factor 5 is Literal Verbatim; Factor 6 is Multiplying; Factor 7 is Relative Gist.

Table 5

Percentage of responses given for the NS in Studies 1 and 2

		Study 1 (*N*=259) %	Study 2 (*N*=190) %
Item 1. Imagine that we roll a fair, six-sided die 1000 times—Out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)?			
Correct response (500)		78.8	91.1
Errors	Matching error (2, 4 or 6)	9.7	2.1
	Place value error (0.5, 5, 50, 5000)	1.2	2.1
	Mindless multiplication of the numbers in the problem: 3×1000 (3000)	1.5	0.0
	Mindless division of the numbers in the problem: 1000/3 (333.33)	1.5	0.5
	Other errors	6.2	4.2
	No response	1.2	0.0
Item 2. In the BIG BUCKS LOTTERY, the chances of winning a $1000 prize are 1%. What is your best guess about how many people would win a $1000?			
Correct response (10)		77.2	88.4
Errors	Place value error (50)	7.7	3.2
	Matching error (1 or 1000)	7.3	6.8
	Other errors	7.3	1.6
	No response	0.4	0.0
Item 3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?			
Correct response (0.1)		59.8	68.9
Errors	Place value error (0.001, 0.01, 10, 100)	20.1	24.2
	Matching error (1 or 1000)	13.5	4.7
	Other errors	4.6	2.1
	No response	1.9	0.0
Item 4. Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10			
Correct response (1 in 10)		92.2	98.4
Errors	1 in 100	4.7	1.1
	1 in 1000	3.1	0.5
Item 5. Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%			
Correct response (10%)		93.0	97.4
Errors	1%	4.7	0.5
	5%	2.3	2.1
Item 6. If Person A's risk of getting a disease is 1% in 10 years, and Person B's risk is double that of A's, what is B's risk?			
Correct response (2%)		80.8	91.1
Errors	Divided the number of years by two instead of multiplying the percentage by two (1% in 5years)	7.0	3.7
	Doubled percentage and number of years (2% in 20years)	2.7	0.0
	Other errors	7.6	5.3
	No response	1.9	0.0
Item 7. If Person A's chance of getting a disease is 1 in 100 in 10years, and Person B's risk is double that of A, what is B's risk?			
Correct response (2 out of 100)		77.3	88.9

		Study 1 (*N*=259) %	**Study 2 (*N*=190) %**
Errors	Divided the number of years by two instead of multiplying the percentage by two (1 out of 100 in 5years)	3.1	2.7
	Mindless division of the numbers in the problem: 100/2 and 10/2 (1 out of 50 in 5years)	2.0	0.5
	Doubled every information in the problem (2 out of 200; 2 out of 200 in 20years)	1.6	0.0
	Other errors	14.8	7.9
	No response	1.2	0.0
Item 8. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?			
Correct response (10)		96.5	97.9
Errors	Place value error (1, 100)	1.5	1.6
	Other errors	1.5	0.5
	No response	0.4	0.0
Item 9. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?			
Correct response (100)		93.4	94.7
Errors	Place value error (1)	2.3	1.1
	Matching error (10 or 1000)	1.5	3.7
	Other errors	2.3	0.5
	No response	0.4	0.0
Item 10. If the chance of getting a disease is 20 out of 100, this would be the same as having a ____% chance of getting the disease			
Correct response (20)		90.3	91.6
Errors	Place value error (0.2, 2)	3.5	1.6
	Divided numbers in the problem: 100/20	1.2	2.6
	Other errors	4.6	4.2
	No response	0.4	0.0
Item 11. The chance of getting a viral infection is 0.0005. Out of 10000 people, about how many of them are expected to get infected?			
Correct response (5)		51.7	81.6
Errors	Place-value error (0.0000005; 0.00005; 0.005; 0.05; 0.5; 50; 500 or 5000)	20.5	9.5
	Other errors	22.8	8.4
	No response	5.0	0.5

Table 6

Percentage of responses given for the CRT in Studies 1 and 2

		Study 1 (*N*=259) %	Study 2 (*N*=190) %
Item 1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? ____ cents			
Correct response (5)		30.9	47.9
Errors	Mindless answer (10)	64.1	38.4
	Other errors	2.7	10.0
	No response	2.3	3.7
Item 2. If it takes 5 machines and 5minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes			
Correct response (5)		37.8	38.4
Errors	Mindless answer (100)	47.5	50.5
	Place value error for the mindless answer (1, 10, 1000)	5.0	3.2
	Mindless multiplication of the numbers in the problem (5×100=500)	4.6	2.1
	Mindless division of the numbers in the problem (100/5=20)	1.9	4.2
	Other errors	1.7	1.5
	No response	1.5	0.5
Item 3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days			
Correct response (47)		41.7	63.7
Errors	Mindless division of the numbers in the problem (48/2=24)	34.7	21.0
	Mindless division of the numbers in the problem: (48/2)/2=(12)	6.9	2.6
	Mindless multiplication of the numbers in the problem (48×2=96)	1.2	0.0
	Other errors	14.0	12.1
	No response	1.5	0.5

Note: Some columns do not add to 100 because of rounding.

Table 7

Study 1. Correlations between test scores, memory scores, fallacies, and factor scores

	General Numeracy	Expanded Numeracy	NS	Cognitive Ability	Preference for Display Numerical Information	SNS	CRT	Memory deviation score	Gist memory score	No. of conjunction fallacies	No. of disjunction fallacies
General Numeracy	1										
Expanded Numeracy	0.38**	1									
NS	0.76**	0.89**	1								
Cognitive Ability	0.42**	0.37**	0.47**	1							
Preference for Display Numerical Information	0.38**	0.21**	0.34**	0.48**	1						
SNS	0.46**	0.34**	0.47**	0.88**	0.84**	1					
CRT	0.55**	0.34**	0.51**	0.48**	0.34**	0.48**	1				
Memory deviation score	−0.28**	−0.20**	−0.28**	−0.14*	−0.18**	−0.18**	−0.27**	1			
Gist memory score	−0.08	−0.00	−0.04	−0.16*	−0.09	−0.14*	−0.20**	0.42**	1		
No. of conjunction fallacies	−0.34**	−0.31**	−0.39**	−0.28**	−0.22**	−0.30**	−0.20**	0.18*	−0.08	1	
No. of disjunction fallacies	−0.28**	−0.13*	−0.23**	−0.24**	−0.14*	−0.23**	−0.26**	0.18**	0.00	0.56**	1
F1 Cognitive Ability	0.13*	0.21**	0.21**	0.92**	0.40**	0.79**	0.28**	−0.04	−0.13*	−0.19**	−0.16*
F2 Verbatim	0.75**	0.26**	0.56**	0.25**	0.20**	0.27**	0.88**	−0.30**	−0.15*	−0.22**	−0.25**
F3 Proportions	0.08	0.59**	0.45**	0.03	0.09	0.07	0.08	−0.15*	0.05	−0.16*	−0.04
F4 Gist	0.17**	0.44**	0.39**	0.10	0.02	0.08	−0.02	−0.03	0	−0.06	−0.03
F5 Multiplying	0.23**	0.48**	0.45**	0.14*	−0.02	0.08	0.03	−0.03	0.09	−0.24**	−0.08
F6 Preference for Display Numerical Information	0.31**	−0.07	0.11	0.10	0.83**	0.51**	0.01	−0.08	−0.01	−0.12	−0.07

Note: General Numeracy subscale score=items 1–3 of NS; Expanded Numeracy subscale score=items 4–11 of NS; NS=Objective Numeracy Scale score; Cognitive Ability subscale score=items 1–4 of SNS; Preference for Display Numerical Information subscale score=items 5–8 of SNS; SNS, Subjective Numeracy Scale score; CRT, Cognitive Reflection Test score; F1=Factor 1 score (Subjective Numeracy's subscale Cognitive Ability); F2= Factor 2 score (Verbatim Monitoring); F3=Factor 3 score (Proportions); F4=Factor 4 score (Relative Magnitude/Gist); F5=Factor 5 score (Multiplying); F6=Factor 6 score (Subjective Numeracy's subscale Preference for Display Numerical Information).

* Correlation is significant at the 0.05 level (two-tailed).

** Correlation is significant at the 0.01 level (two-tailed).

Table 8

Study 2. Correlations between test scores, bias, and factor scores

	General numeracy	Expanded numeracy	NS	Cognitive ability	Preference for Display Numerical Information	SNS	CRT	APM	Need for cognition	Faith in Intuition	Go/no-go reaction time	Go/no-go correct answers	Go/no-go false alarms	Ratio bias
General Numeracy	1													
Expanded Numeracy	0.38**	1												
NS	0.82**	0.84**	1											
Cognitive Ability	0.44**	0.29**	0.44**	1										
Preference for Display Numerical Information	0.36**	0.18*	0.32**	0.47**	1									
SNS	0.47**	0.28**	0.45**	0.87**	0.84**	1								
CRT	0.37**	0.29**	0.39**	0.41**	0.33**	0.43**	1							
APM	0.27**	0.19**	0.28**	0.24**	0.22**	0.27**	0.31**	1						
Need for Cognition	0.32**	0.15*	0.28**	0.52**	0.41**	0.54**	0.33**	0.29**	1					
Faith in Intuition	−0.04	−0.06	−0.06	0.04	−0.07	−0.01	−0.08	−0.15*	0.03	1				
Go/no-go reaction time	−0.08	−0.03	−0.06	−0.14	−0.12	−0.15*	−0.07	−0.02	−0.13	−0.01	1			
Go/no-go correct answers	0.00	0.08	0.05	0.09	0.04	0.07	0.03	0.05	0.08	0.01	−0.63**	1		
Go/no-go false alarms	−.09	−0.02	−0.06	−0.07	−0.02	−0.06	−0.02	−0.19**	−0.05	0.03	−0.52**	0.12	1	
Ratio bias	0.09	0.22**	0.19**	0.09	0.04	0.08	0.13	0.21**	0.03	−0.09	0.04	−0.01	−0.13	1
F1 Cognitive Ability	0.29**	0.12	0.25**	0.93**	0.48**	0.84**	0.24**	0.19**	0.53**	0.03	−0.16*	0.08	−0.02	0.03
F2 Proportions	0.38**	0.58**	0.58**	0.15*	−0.15*	0.01	0.05	0.09	−0.02	−0.04	0.08	−0.01	−0.07	0.25**
F3 CRT	0.24**	0.14	0.22**	0.19**	0.05	0.15*	0.91**	0.23**	0.19**	−0.02	−0.04	0.00	−0.02	0.09
F4 Preference for Display Numerical Information	0.40**	0.01	0.24**	0.08	0.73**	0.45**	0.11	0.12	0.14	−0.06	−0.01	−0.03	−0.09	0.03
F5 Verbatim	0.42**	0.49**	0.55**	0.08	0.19**	0.16*	0.19**	0.13	0.08	−0.13	−0.05	0.02	0.00	−0.01
F6 Multiplying	−0.03	0.37**	0.21**	0.02	0.19**	0.12	0.02	−0.01	0.07	0.05	−0.07	0.03	0.17*	0.01
F7 Gist	−0.09	0.19**	0.07	−0.07	0.20**	0.07	0.10	0.21**	0.16*	0.00	0.11	−0.01	−0.04	0.18*

Note: General Numeracy subscale score=items 1–3 of NS; Expanded Numeracy subscale score=items 4–11 of NS; NS=Objective Numeracy scale score; Cognitive Ability subscale score=items 1–4 of SNS; Preference for Display Numerical Information subscale score=items 5–8 of SNS; SNS, Subjective Numeracy Scale score; CRT, Cognitive Reflection Test score; APM, Advanced Progressive Matrices; go/no-go reaction time=mean; Factor 1 is Subjective Numeracy Cognitive Abilities; Factor 2 is Proportions; Factor 3 is Cognitive Reflection Test; Factor 4 is SNS Preference for Display Numerical Information; Factor 5 is Literal Verbatim; Factor 6 is Multiplying; Factor 7 is Relative Gist.

* Correlation is significant at the 0.05 level (two-tailed).

** Correlation is significant at the 0.01 level (two-tailed).

Table 9

Study 1. Factor loadings for factor analysis with varimax rotation for NS and CRT

	Component			
Items for each scale	**1**	**2**	**3**	**4**
CRT 1—A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?	**0.73**	−0.04	0.09	−0.02
CRT 2—If it takes five machines 5minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?	**0.74**	0.10	−0.09	0.03
CRT 3—In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?	**0.76**	0.19	−0.05	0.10
NS 1—Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)?	**0.56**	0.11	0.26	0.17
NS 2—In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS?	**0.52**	0.16	0.14	0.20
NS 3—In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?	**0.64**	−0.01	0.04	0.15
NS 4—Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10	0.10	0.09	**0.89**	0.02
NS 5—Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%	0.02	0.12	**0.87**	0.00
NS 6—If Person A’s risk of getting a disease is 1% in 10years, and Person B’s risk is double that of A’s, what is B’s risk?	0.14	0.10	0.01	**0.83**
NS 7—If Person A’s chance of getting a disease is 1 in 100 in 10years, and Person B’s risk is double that of A, what is B’s risk?	0.22	0.08	0.03	**0.83**
NS 8—If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?	0.05	**0.81**	0.24	0.17
NS 9—If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?	−0.04	**0.83**	0.06	0.14
NS 10—If the chance of getting a disease is 20 out of 100, this would be the same as having a ____% chance of getting the disease.	0.32	0.45	0.31	0.15
NS 11—The chance of getting a viral infection is 0.0005. Out of 10000 people, about how many of them are expected to get infected?	0.25	0.48	−0.09	−0.21
Eigenvalues	2.91	1.90	1.82	1.60

Note: Factor loadings >0.50 are in boldface. CRT, Cognitive Reflection Test; NS, Objective Numeracy Scale; Factor 1 is Verbatim/Monitoring; Factor 2 is Proportions; Factor 3 is Relative Magnitude/Gist, and Factor 4 is Multiplying.

Table 10

Study 1. Predictors of memory deviation and judgment fallacies

	Unstandardized coefficients		Standardized coefficients		
	B	**SE**	**Beta**	***t***	**Sig.**
Dependent variable: memory deviation					
(Constant)	35.07	1.16		3.20	<0.001
Factor 1=Verbatim/Monitoring	−5.85	1.16	−0.30	−5.03	<0.001
Factor 2=Proportions	−3.14	1.16	−0.16	−2.70	0.01
Factor 3=Relative Magnitude/Gist	−0.69	1.16	−0.03	−0.59	0.55
Factor 4=Multiplying	−0.51	1.16	−0.03	−0.44	0.66
Dependent variable: gist memory score					
(Constant)	6.96	0.54		12.95	<0.001
Factor 1=Verbatim/Monitoring	−1.64	0.54	−0.19	−3.05	0.00
Factor 2=Proportions	0.20	0.54	0.02	0.38	0.71
Factor 3=Relative Magnitude/Gist	−0.07	0.54	−0.01	−0.13	0.89
Factor 4=Multiplying	0.67	0.54	0.08	1.25	0.21
Dependent variable: number of conjunction fallacies					
(Constant)	10.17	0.34		29.79	<0.001
Factor 1=Verbatim/Monitoring	−1.35	0.35	−0.26	−3.85	<0.001
Factor 2=Proportions	−0.74	0.30	−0.16	−2.45	0.02
Factor 3=Relative Magnitude/Gist	−0.29	0.31	−0.06	−0.94	0.35
Factor 4=Multiplying	−1.14	0.32	−0.24	−3.60	<0.001
Dependent variable: number of disjunction fallacies					
(Constant)	11.21	0.31		35.70	<0.001
Factor 1=Verbatim/Monitoring	−1.51	0.31	−0.29	−4.80	<0.001
Factor 2=Proportions	−0.18	0.31	−0.03	−0.56	0.58
Factor 3=Relative Magnitude/Gist	−0.21	0.31	−0.04	−0.67	0.50
Factor 4=Multiplying	−0.47	0.31	−0.09	−1.50	0.14

SE, standard error.

Table 11

Study 2. Factor loadings for factor analysis with varimax rotation for NS, SNS, CRT, APM, NFC, FI, and go/no-go scores

	Component								
Item for each scale	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
CRT 1—A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?	0.11	−0.02	0.00	**0.80**	0.04	− 0.02	0.03	− 0.10	0.13
CRT 2—If it takes 5 machines 5minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?	0.18	−0.01	0.01	**0.68**	−0.05	0.17	− 0.01	0.26	−0.09
CRT 3—In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?	0.34	0.15	0.03	**0.54**	0.03	0.16	0.14	0.27	0.07
NS 1—Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)?	0.20	**0.56**	−0.05	0.21	−0.03	0.13	0.38	− 0.10	−0.09
NS 2—In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS?	0.22	0.37	−0.02	0.24	0.05	− 0.20	0.34	0.30	0.09
NS 3—In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?	0.32	0.06	0.05	0.16	0.03	0.11	0.01	**0.54**	0.18
NS 4—Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10	−0.01	0.38	0.02	0.12	0.04	**0.69**	0.07	− 0.07	0.04
NS 5—\Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%	0.15	−0.16	−0.12	0.00	−0.03	**0.69**	− 0.06	− 0.02	−0.09
NS 6—If Person A’s risk of getting a disease is 1% in 10years, and Person B’s risk is double that of A’s, what is B’s risk?	0.06	−0.03	0.03	− 0.13	**0.82**	− 0.01	0.11	0.07	−0.01
NS 7—If Person A’s chance of getting a disease is 1 in 100 in 10years, and Person B’s risk is double that of A, what is B’s risk?	0.09	0.21	−0.01	0.16	**0.75**	−0.02	0.11	0.11	−0.14
NS 8—If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?	0.00	**0.71**	−0.04	− 0.07	−0.03	0.16	− 0.18	0.13	0.30
NS 9—If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?	0.27	0.47	0.10	0.02	0.07	0.03	− 0.20	0.11	**0.55**
NS 10—If the chance of getting a disease is 20 out of 100, this would be the same as having ____a % chance of getting the disease.	0.04	**0.66**	−0.03	− 0.04	0.12	− 0.05	− 0.04	0.07	−0.11
NS 11—The chance of getting a viral infection is 0.0005. Out of 10000 people, about how many of them are expected to get infected?	0.02	0.08	−0.04	0.05	0.11	− 0.06	0.00	**0.77**	−0.01
SNS–CA 1—How good are you at working with fractions?	**0.72**	0.09	0.07	0.15	0.01	0.07	− 0.08	0.25	0.03
SNS–CA 2—How good are you at working with percentages?	**0.86**	0.04	−0.01	− 0.03	−0.07	0.09	0.06	0.09	0.06
SNS–CA 3—How good are you at calculating a 15% tip?	**0.69**	0.26	0.11	0.17	−0.12	− 0.21	0.09	0.04	−0.14
SNS–CA 4—How good are you at figuring out how much a shirt will cost if it is 25% off?	**0.75**	0.24	−0.04	0.17	0.10	− 0.13	0.09	− 0.03	−0.16
SNS–PDNI 1—When reading the newspaper, how helpful do you find tables and graphs that are parts of a story?	**0.52**	−0.23	0.07	− 0.03	0.18	0.15	0.23	0.20	0.27
SNS–PDNI 2—When people tell you the chance of something happening, do you prefer that they use words (“it rarely happens”) or numbers (“there’s a 1% chance”)?	0.40	−0.02	0.05	0.20	0.20	0.03	0.48	− 0.20	0.08
SNS–PDNI 3—When you hear a weather forecast, do you prefer predictions using percentages (e.g., “there will be a 20% chance of rain today”) or predictions using only words (e.g., “there is a small chance of rain today”)?	0.09	−0.09	0.02	− 0.02	0.13	0.05	**0.74**	0.07	0.10

	Component								
Item for each scale	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
SNS–PDNI 4—How often do you find numerical information to be useful?	**0.66**	−0.02	−0.03	0.12	0.19	0.21	0.17	− 0.06	0.21
Advanced Progressive Matrices score	0.17	0.06	0.06	0.24	−0.11	0.47	0.24	0.25	0.07
Need-for-cognition score	**0.65**	−0.06	0.07	0.16	0.07	0.27	0.07	0.07	−0.09
Faith-in-Intuition score	0.08	0.02	0.05	− 0.08	0.15	0.06	− 0.22	− 0.06	−**0.72**
Go/no go—reaction time	−0.13	0.07	−**0.90**	− 0.08	−0.09	0.10	0.03	0.03	−0.06
Go/no go—correct answers	0.02	0.04	**0.86**	− 0.08	−0.10	0.06	0.11	0.07	−0.11
Go/no go—false alarms	−0.03	−0.12	0.48	0.13	0.38	− 0.15	− 0.34	− 0.24	0.25
Eigenvalues	4.09	2.09	1.86	1.84	1.64	1.59	1.51	1.48	1.31

Note: Factor loadings >0.50 are in boldface. CRT, Cognitive Reflection Test; NS, Objective Numeracy scale; SNS, Subjective Numeracy Scale; CA, Subjective Numeracy's subscale Cognitive Ability; PDNI, Subjective Numeracy's subscale Preference for Display Numerical Information.

Table 12

Study 2. Factor loadings for factor analysis with varimax rotation for NS and CRT

	Component				
Items for each scale	**1**	**2**	**3**	**4**	**5**
CRT 1—A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?	−0.01	**0.76**	− 0.01	− 0.09	−0.03
CRT 2—If it takes 5 machines 5minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?	−0.05	**0.70**	− 0.02	0.20	0.14
CRT 3—In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?	0.14	**0.65**	0.07	0.35	0.17
NS 1—Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)?	**0.54**	0.39	0.09	− 0.10	0.15
NS 2—In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS?	0.40	0.42	0.13	0.25	−0.22
NS 3—In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?	0.11	0.27	0.00	**0.63**	0.03
NS 4—Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10	0.38	0.11	0.02	− 0.02	**0.67**
NS 5—Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%	−0.16	0.07	−0.03	0.05	**0.83**
NS 6—If Person A’s risk of getting a disease is 1% in 10years, and Person B’s risk is double that of A’s, what is B’s risk?	−0.03	−0.11	**0.86**	0.07	−0.01
NS 7—If Person A’s chance of getting a disease is 1 in 100 in 10years, and Person B’s risk is double that of A, what is B’s risk?	0.15	0.16	**0.84**	0.05	−0.01
NS 8—If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?	**0.76**	−0.12	− 0.11	0.22	0.11
NS 9—If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?	**0.56**	0.02	− 0.02	0.42	0.02
NS 10—If the chance of getting a disease is 20 out of 100, this would be the same as having a ____ %chance of getting the disease.	**0.66**	0.03	0.15	− 0.09	−0.07
NS 11—The chance of getting a viral infection is 0.0005. Out of 10000 people, about how many of them are expected to get infected?	0.00	0.02	0.10	**0.77**	−0.01
Eigenvalues	2.01	1.95	1.51	1.48	1.28

Note: Factor loadings >0.50 are in boldface. CRT, Cognitive Reflection Test; NS, Objective Numeracy Scale; Factor 1 is Proportions; Factor 2 is Monitoring (Cognitive Reflection Test); Factor 3 is Multiplying; Factor 4 is Verbatim Matching, and Factor 5 is Relative Magnitude/Gist.

Table 13

Study 2. Predictors of ratio bias

Ratio bias (correct/wrong)	**B**	**SE**	**Wald**	**d.f.**	**Sig.**	**Exp(B)**
Factor 1=Proportions	0.57	0.21	7.23	1.00	0.01	1.77
Factor 2=Monitoring (Cognitive Reflection Test)	0.20	0.18	1.32	1.00	0.25	1.22
Factor 3=Multiplying	0.03	0.18	0.04	1.00	0.85	1.04
Factor 4=Verbatim Matching	0.04	0.18	0.06	1.00	0.80	1.05
Factor 5=Relative Magnitude/ Gist	0.50	0.21	5.89	1.00	0.02	1.65
Constant	1.24	0.18	46.29	1.00	0.00	3.47

SE, standard error.