Cognitive Psychology 118 (2020) 101273

Contents lists available at ScienceDirect

Cognitive Psychology

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

Intuitive symbolic magnitude judgments and decision making under risk in adults

Andrea L. Patalano[a,*], Alexandra Zax[a], Katherine Williams[a], Liana Mathias[a], Sara Cordes[b], Hilary Barth[a]

[a] *Department of Psychology, Wesleyan University, United States*
[b] *Department of Psychology, Boston College, United States*

ARTICLE INFO

Keywords:
Decision making under risk
Cumulative prospect theory
Probability weighting
Numeracy
Proportion judgment
Number line estimation

ABSTRACT

Performance on an intuitive symbolic number skills task—namely the number line estimation task—has previously been found to predict value function curvature in decision making under risk, using a cumulative prospect theory (CPT) model. However there has been no evidence of a similar relationship with the probability weighting function. This is surprising given that both number line estimation and probability weighting can be construed as involving proportion judgment, that is, involving estimating a number on a bounded scale based on its proportional relationship to the whole. In the present work, we re-evaluated the relationship between number line estimation and probability weighting through the lens of proportion judgment. Using a CPT model with a two-parameter probability weighting function, we found a double dissociation: number line estimation bias predicted probability weighting curvature while performance on a different number skills task, number comparison, predicted probability weighting elevation. Interestingly, while degree of bias was correlated across tasks, the direction of bias was not. The findings provide support for proportion judgment as a plausible account of the shape of the probability weighting function, and suggest directions for future work.

1. Introduction

1.1. Decision making under risk

Imagine having a choice between an option that will give you a sure gain of $100 and an option that will give you a 50% chance of gaining $300 otherwise nothing. Which do you prefer? How did you arrive at this preference? According to cumulative prospect theory (CPT; Tversky & Kahneman, 1992), a dominant model of decision making under risk, an overall valuation of each option is made, and the option with the higher value is selected. In the model, outcomes are transformed into *subjective values* and probabilities are transformed into *decision weights* prior to integration. The subjective value transformation reflects that outcomes are increasingly underestimated as their magnitude increases, following a concave curve (see Fig. 1a, which shows the function for gains only; Tversky & Kahneman, 1992). The probability weighting transformation reflects that small probabilities are typically overestimated and large ones underestimated following an inverse S-shaped curve (see Fig. 1b), with a crossover at around 0.4. These aggregate probability weighting patterns notwithstanding, individual-level variation is quite large (Gonzalez & Wu, 1999; Patalano, Saltiel, Machlin, & Barth, 2015) and relatively stable (Birnbaum, 1974; Glöckner & Pachur, 2012; Zeisberger, Vrecko, & Langer, 2012), and

* Corresponding author at: Department of Psychology, Wesleyan University, 207 High Street, Middletown, CT 06459, United States.
E-mail address: apatalano@wesleyan.edu (A.L. Patalano).

https://doi.org/10.1016/j.cogpsych.2020.101273
Received 30 March 2019; Received in revised form 30 October 2019; Accepted 7 January 2020
Available online 03 February 2020

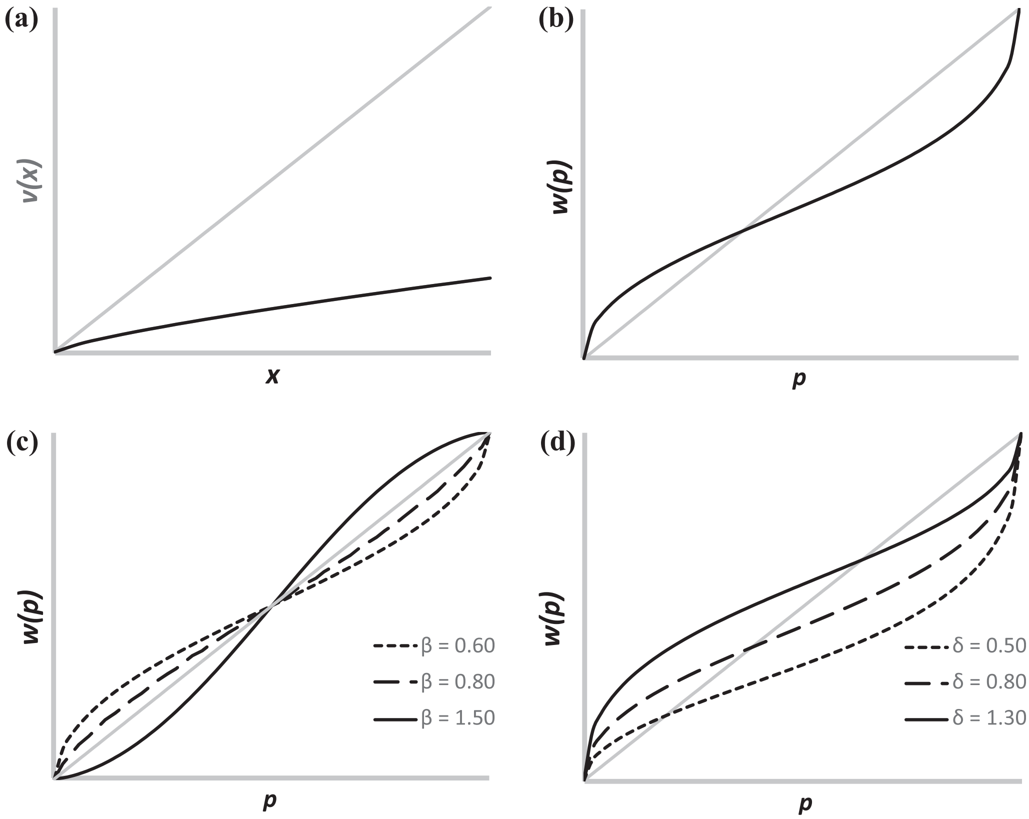

Fig. 1. (a) Value function for gains, $v(x) = x^{\alpha}$ with $\alpha = 0.77$ (black curve). (b) Probability weighting function $w(p) = \delta^{\beta}/(\delta p^{\beta} + (1 - p)^{\beta})$ with δ (elevation) = 0.77, and β (curvature) = 0.53 (black curve; medians from Patalano et al., 2015). (c) Result of changing probability weighting curvature β while holding probability weighting elevation δ constant. (d) Result of changing probability weighting elevation δ while holding probability weighting curvature β constant.

can differ qualitatively from aggregate patterns such that ~25% of participants may have an S-shaped rather than inverse S-shaped curve (Patalano et al., 2015).

Two important questions for understanding probability weighting are of particular concern to us. Why do these probability weighting curves take the shapes that they do? And, relatedly, why are there such large individual differences in degree and direction of curvature? A number of psychological explanations for aggregate-level value and probability weighting curves have been proposed including diminishing sensitivity to probabilities farther from certainty (Tversky & Kahneman, 1992), anchoring plus adjustment through mental simulation (Venture Theory; Hogarth & Einhorn, 1990), an affect-based account in which hope and fear influence judgments (Rottenstreich & Hsee, 2001), and dynamic ranking of probabilities relative to others sampled from the decision environment (Decision by Sampling Theory; Stewart, 2009, Stewart, Chater, & Brown, 2006, Stewart, Reimers, & Harris, 2015; but see Alempaki et al., 2019). These explanations generally assume that patterns of bias arise from inputs and mental processes specific to decision making, rather than from more general characteristics of cognition. In the present work, we consider whether conceptualizing probability use in decision making as a task of proportion judgment, and bringing to bear a proportion judgment model, can offer insight into probability weighting curvature. We first review the broader literature on the relationship between numerical cognition and more veridical use of numbers in decision making; we then present a proportion judgment framework; and we motivate and introduce the present study.

1.2. Decision making and number skills

There is a large literature on the relationship between numeracy, defined as mathematical quantitative literacy (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012), and effective decision making. It is well documented that numeracy, often assessed with a brief scale such as one involving translating between part-whole formats (e.g., fractions, percentages, and decimals; Lipkus, Samsa, & Rimer, 2001), is associated with greater (and more effective) use of number-based decision strategies (e.g., as opposed to use of narratives, affect, or intuition; see Peters, 2012; Reyna, Nelson, Han, & Dieckmann, 2009, for review). Numeracy has also been associated specifically with more veridical use of decision-related numbers (i.e., Black, Nease, & Tosteson, 1995; Patalano, Lolli, & Sanislow, 2018; Winman, Juslin, Lindskog, Nilsson, & Kerimi, 2014; Peters, 2012). Most directly relevant to the present work, in a hypothetical gambling task from which cumulative prospect model curves could be inferred, numeracy was associated with a more linear value function (closer to the identity line; Patalano et al., 2015; Schley & Peters, 2014). However, findings remain mixed regarding the probability weighting function, with one study reporting a moderate correlation (Patalano et al., 2015) but not another (Schley & Peters, 2014). Given the diversity of skills associated with numeracy which can range from number operations and computation to simply having more positive affective responses to working with numbers (Peters & Bjalkebring, 2015) the findings do not point to specific mechanisms that might underlie curvature, or *bias*, resulting from the use of numbers in decision making.

In many accounts of the psychological representation of number, symbolic numbers are thought to lie on an internal continuum of mental magnitudes (Dehaene, Dupoux, & Mehler, 1990; Gallistel & Gelman, 1992). Numbers are mapped to magnitudes that are represented in an approximate fashion. Some evidence of this comes from *number comparison tasks* (Moyer & Landauer, 1967; see also Dehaene et al., 1990) in which one must rapidly decide which of two numbers (e.g., 47 vs. 59) is larger. Longer response times are observed for numbers closer to one another in a pair, called a distance effect (3 vs. 4 takes longer than 3 vs. 9), and for pairs of greater magnitude, called a size effect (1 vs. 3 is faster than 7 vs. 9; Parkman, 1971), the latter suggesting decreasing discriminability of larger magnitudes. Based on effects such as these, mental magnitudes are often described as a series of overlapping normal distributions (e.g., Dehaene, 1997; Gallistel & Gelman, 2005), with distribution means getting closer together (Dehaene, 1992, 2003) or variances getting larger (Gallistel & Gelman, 1992; Meck, Church, & Gibbon, 1985) as magnitude increases. To our knowledge, Peters and colleagues (Peters, Slovic, Västfjäll, & Mertz, 2008) were the first to propose and test the idea that these approximate mental representations might underlie the transformation of numbers to the internal magnitudes that are used in decision making.

Performance on a different numerical task, the *number line estimation task*, has also been related to decision making. This task calls for precise localization of a magnitude on a scale (e.g., rather than simply an ordinal judgment as is the case with number comparison). In a typical version of the task (Barth & Paladino, 2011; Siegler & Opfer, 2003) a number is flashed on a screen (e.g., 750) followed by a line labeled only with endpoints (e.g., 0 and 1000), and the task is to indicate on the line where the target number belongs. Error in number placement (based on the difference between the selected location and the actual one) is one commonly used individual difference measure of task performance. Schley and Peters (2014) had participants complete a 16-question forced-choice gambling task (e.g., Would you rather have a 50% chance of $100 or a 75% chance of $80?; from Toubia, Johnson, Evgeniou, & Delquié, 2013), in which stimuli were dynamically constructed based on one's prior choices for efficient estimation of parameters. Number line estimation error was correlated with cumulative prospect model value function curvature ($r \approx 0.35$ across studies) but not with probability weighting curvature (Schley & Peters, 2014; see also Peters & Bjalkebring, 2015). The authors argued that the curvilinear error pattern in the value function is similar to number line estimation bias and, like the size effect seen earlier, reflects a declining discrimination sensitivity as magnitude increases.

As indicated earlier, we are interested in the shape of the probability weighting function in decision making under risk. To date, other than one numeracy finding, there is surprisingly little evidence that number skills are generally related to linearity of the probability weighting function. Further, based on existing work, one might conclude that there is little evidence of an inverse S-shaped or S-shaped bias in any numerical tasks. However, for the number line estimation task, there is actually growing consensus that if one models the relationship between actual and estimated magnitude, the best fitting curve is inverse S-shaped or S-shaped (similar to the probability weighting curve), rather than a concave curve (see, e.g., Ashcraft & Moore, 2012; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Rouder & Geary, 2014; Slusser & Barth, 2017; see also Sullivan, Juhasz, Slattery, & Barth, 2011; but see Siegler & Opfer, 2003, for an alternative perspective). It may be fruitful, then, to consider models of number line estimation bias as a potential avenue to understanding the source of the inverse S-shaped curve in probability weighting patterns and individual differences in degree of curvature.

1.3. Proportion judgment framework

The S-shaped and inverse S-shaped curves, rather than being unique to decision making, have been found in a wide variety of cognitive, perceptual, and motor tasks with the common property that they can be conceptualized as proportion judgment (for reviews, see Hollands & Dyre, 2000; Zhang & Maloney, 2012). For example, when individuals estimate the proportion of black dots to the total number of black and white dots in a display, they tend to overestimate small proportions and to underestimate larger ones, following an inverse S-shaped curve (Varey, Mellers, & Birnbaum, 1990). The same is true in the auditory domain when participants are asked to judge the ratio of pairs of silent time intervals bounded by clicks (Nakajima, 1987), or even when people estimate from memory the proportion of individuals in society who belong to various demographic categories (see Landy, Guay, & Marghetis, 2018). There are also situations in which the pattern is S-shaped rather than inverse S-shaped at the group level, such as when Shuford (1961) showed arrays of red and blue squares and participants estimated the percentage of a given color, or, in Simpson and Voss (1961), when participants estimated the percentage of light flashes over time.

A model was developed by Hollands and Dyre (2000), called the cyclical power model, to explain bias in perceptual tasks. The model builds on Stevens' Law (Stevens, 1957) which describes the psychophysical relationship between the estimated or perceived magnitude of a physical stimulus (e.g., brightness) and its actual magnitude as a power function $y = \delta x^{\beta}$, where β quantifies bias associated with judgments of a stimulus continuum (and δ is a scaling parameter). A value of $\beta < 1$ produces a concave curve and $\beta > 1$ produces a convex curve. Spence (1990) extended Steven's Law to proportion judgments by showing that when observers respond based on two bounding reference points (0 and 1, where 1 refers to the whole) in their judgments, estimates are predicted by $y = x^{\beta}/(x^{\beta} + (1 - x)^{\beta})$, with the value of β determining magnitude and direction of bias. This function produces an inverse S-shaped curve when $\beta < 1$ (i.e., when the original power function is concave), and an S-shaped curve when the $\beta > 1$ (i.e., when the original power function is convex). (Note that this equation is the same as the curvature component of the probability weighting function, which will be described later.) Hollands and Dyre (2000) generalized Spence's model to explain the wide range of estimation

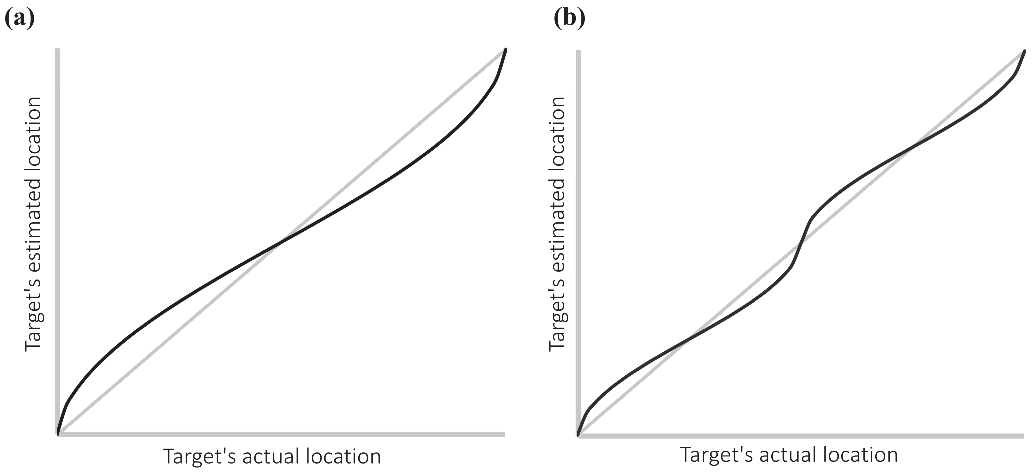

Fig. 2. (a) Example of a cyclical power model (CPM) curve following a one-cycle pattern: $y = x^{\beta}/(x^{\beta} + (1 - x)^{\beta})$; (b) Curve for two-cycle pattern with same bias ($\beta = 0.65$ in both images). A $\beta > 1$ would produce an S-shaped curve instead of the inverse S-shaped curves shown here.

patterns—such as the one- versus two-cycle curves shown in Fig. 2—that can arise from using additional reference points (such as a midpoint) during proportion judgment.

The Hollands and Dyre (2000) model has been extended from perceptual stimuli to symbolic numbers such as those used in the number line estimation task. Like judging the relative numbers of black dots in a larger set, placing 650 on a number line bounded by 0 and 1000 also involves judging the relationship of a part to a whole, and thus can be construed as proportion judgment. When adults place numbers on a bounded number line, error is not unsystematic: rather, the bias in response can be well-described by inverse S-shaped or S-shaped curve, with one or two cycles, with the latter arising from use of a midpoint as an additional reference point (e.g., Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Slusser, Santiago, & Barth, 2013). This pattern is not surprising given that symbolic numbers have been linked to approximate representations of magnitude, and that the task can be straightforwardly construed as proportion estimation. Estimation data have also been collected using an unbounded number line (in which only the left side of the number line is labeled along with the size of a single unit), a task that has been proposed to better assess magnitude estimation (as modeled with Steven's Law) rather than proportion judgment (Cohen & Blanc-Goldhammer, 2011; Cohen, Blanc-Goldhammer, & Quinlan, 2018). In one study, group-level bias was found to be about the same in bounded and unbounded tasks ($\beta = \sim 1.12$; Cohen et al., 2018), consistent with the shape of the curve in the bounded task arising from proportional relationship between approximate magnitudes.

In the same way that estimating 650 on a 0–1000 scale can be construed as a proportion judgment task, so can interpreting a probability such as 75% in decision making, which involves judging where 75 is located on a 0 to 100 scale. We propose here (see also Patalano et al., 2015) that the bias in proportion judgment in decision making might arise for the same reasons as in the number line estimation task, namely, the use of approximate mental magnitudes in the context of a proportion judgment. Suggestive evidence comes from a study by Müller-Trede, Sher, and McKenzie (2018). When they used an unbounded scale (a scale with no upper bound) to describe different levels of uncertainty, they found diminishing sensitivity to probabilities rather than an inverse S-shaped pattern, suggesting that the latter pattern is specific to the use of a bounded probability scale. There is also some evidence that bias may also reflect skill in proportional reasoning itself in that, across development, a reduction in bias in number line estimation is more strongly related to proportion judgment skill than to a change in magnitude estimation (Cohen & Sarnecka, 2014). In the present study, we predict that if the probability weighting function curve arises from proportion judgment broadly speaking (rather than from, e.g., affective responses, decision-making specific anchoring plus adjustment, etc.), then one's degree of bias in the decision task probability weighting function and the number line estimation task should be related.

Findings from the number line estimation task (and from several perceptual tasks, e.g., Shuford, 1961) are inconsistent with the predictions of Spence's model in at least one regard: *direction* of bias on proportion judgment tasks cannot always be predicted from underlying magnitude estimation. For example, while average bias for the unbounded number line estimation task is consistently a convex curve (i.e., $\beta > 1$; Cohen & Sarnecka, 2014), the bias for the bounded number line estimation task changes across development. Specifically, the pattern changes from inverse S-shaped to S-shaped at ~9 years of age and the latter extends into adulthood (Cohen & Sarnecka, 2014; Slusser & Barth, 2017; Slusser et al., 2013). It may ultimately be possible to accommodate these findings within the cyclical power model, as the shape of the curve is attributed to both imprecise magnitude estimation and proportion judgment strategy. For example, in number line estimation, if individuals differ in whether they treat the distance between the target magnitude and the upper boundary versus the target magnitude and the lower boundary as the part being compared to the greater whole, they should have different response patterns. While we are interested in the direction of curvature across tasks, we do not have reason to predict that the direction (S-shaped vs. inverse S-shaped) will be similar in probability weighting curvature and number line estimation bias.

1.4. Overview of present study

The goal of the present study was to re-evaluate the relationship between probability weighting and number line estimation through the lens of a proportion judgment account. We considered whether we might be able to align models and measures of proportion judgment across tasks, and whether such alignment might reveal any previously obscured relationships between performance. In a first session of the present study, we administered a set of number skills tasks to undergraduate participants. We gave a number line estimation task where we fit the cyclical power model to each participant's data (choosing the best fitting model—one-cycle or two-cycle—for each participant; Slusser & Barth, 2017). From the modeling, we obtained β for each participant, which we used to compute a measure of bias without regard to direction, and also a measure of direction (i.e., S-shaped or inverse S-shaped). Additionally, we administered a second number skills task, namely, the number comparison task for contrast. We fit a distance model (Dehaene et al., 1990; see Methods) to the response time data and computed the estimated slope for each participant as a measure of discrimination difficulty. This task is not a proportion judgment task and performance is not consistently correlated with number line estimation in adults (see Schneider, Thompson, & Rittle-Johnson, 2017, for review). Including the task allowed us to assess the specificity of the relationship between number line estimation bias and probability weighting curvature. Though not related to our central goals, we also administered two numeracy scales, namely, the Lipkus et al. (2001) numeracy scale and the Frederick (2005) cognitive reflection test (see Sinayev & Peters, 2015, for use of this scale as a numeracy measure; see also Weller et al., 2013) and obtained SAT-Math scores from students whose scores were available through the university. Given that standardized test scores are not required for university admission, we did not plan to use this measure in most analyses (as the *n* would be considerably reduced) but we included the scores in pairwise correlations as an additional numeracy measure.

In the second session of the study, we administered a 176-trial hypothetical gambling task using a *certainty equivalent* (*CE*) procedure (based on Gonzalez & Wu, 1999; see Method for details). Fitting the cumulative prospect theory model to each participant's certainty equivalent data, we obtained value curvature (α), probability weighting elevation (δ), and probability weighting curvature (β) parameter estimates for each individual. From each estimate, we computed a measure of absolute deviation from the identity line, where a larger value reflects greater deviation without regard to direction (see Patalano et al., 2015). A separate variable coded the direction of each curve (e.g., inverse S-shaped vs. S-shaped for probability weighting curvature). Note that in the cumulative prospect model equation that we used (Gonzalez & Wu, 1999; see also Fox & Poldrack, 2014), the probability weighting function is the same equation as Spence's equation except that in addition to the curvature parameter β (Fig. 1c), the second parameter δ (an elevation parameter; Fig. 1d) raises or lowers the entire curve. The curvature parameter is often interpreted as reflecting discrimination of probabilities and the elevation parameter is interpreted as the attractiveness of gambling to the individual (Gonzalez & Wu, 1999). Going forward, when we refer to CPT value curvature, probability weighting curvature, and probability weighting elevation, we generally mean bias without regard to direction (i.e., that which is represented by the deviations rather than by the parameter estimates).

We predicted that if number skills related to proportion judgment underlie both number line estimation bias and gambling probability weighting curvature then individual-level bias measures across tasks should be related. We predicted that there would not be a similar relationship between probability weighting curvature and number comparison slope because the latter cannot be readily construed as proportion judgment (or even as magnitude estimation), or between number line estimation and other gambling measures. In past work, numeracy no longer predicted gambling biases once number line estimation was also included in a regression model; Schley & Peters, 2014); that is, the latter explained the relationship between numeracy and gambling biases. In the present work, we use correlations to assess relationships between predictor measures and gambling performance. We additionally use regression analyses to assess whether number line estimation is best described as complementing versus explaining any relationship between numeracy and probability weighting curvature.

2. Method

2.1. Participants

Participants were 91 undergraduate students (68 women; 18–22 years old), who received either introductory psychology course credit or monetary compensation in exchange for their participation. All participants gave written informed consent to participate in the study. Demographic information collected in a prescreening measure included handedness (83 right-handed, 9 left-handed), vision (84 normal or corrected to normal, 6 not normal, 2 not reported), primary language (80 English, 8 another language, 3 not reported), and birthplace (72 United States, 17 another country, 2 not reported). Permission was received from all participants to obtain any standardized achievement test score on record with the University. The *N* was based on convenience (the number of students available from the participant pool), but with the goal of exceeding 62 needed based on a power analysis ($r = 0.35$ $\alpha = 0.05$, $\beta = 0.20$) for observing moderate correlations between number line estimation and gambling task measures.

2.2. Procedure

Participants were run individually in the lab in two sessions. All predictor measures were completed in the first session, while the gambling task was administered in the second session. The two sessions occurred at least two days apart ($M = 13.9$, range = 2–43 days). The first session took approximately 45 min and the second approximately 75 min. Participants completed (in this order): the number line estimation task, the number comparison task, two numeracy-related scales, and the hypothetical gambling task. At the start of the first session, they also completed a spatial task, and at the end of the second session, several personality-related measures, which are unrelated to the present report. SAT scores (treated here as another numeracy-related measure) were obtained after data collection was completed.

2.3. Number tasks

2.3.1. Number line estimation task

This task (also described in Barth, Lesser, Taggart, & Slusser, 2015) assesses accuracy in placing a number on a bounded scale. Stimuli were displayed in MATLAB in a different pseudorandom order for each participant. Each trial consisted of a centered fixation rectangle (grey, 12.3 cm × 0.7 cm; for 500 ms) immediately followed by a stimulus screen (for 500 ms) and then a response screen (for 1500 ms). The stimulus screen presented each target value as a numeral (e.g., ‘47′ appeared on the screen). The response screen displayed a 12.3 cm horizontal line with end-points (short vertical lines at each end extending 0.3 cm above and below the line) labeled with ‘0′ and ‘1000′ respectively. The response line appeared in a different pseudorandom location for each trial. We used a standard version of the task with whole-number targets distributed over the entire number-line range, rather than the less common version (with decimal targets and values focused on lower end of scale) used in Schley and Peters (2014) work. Target values were sampled at intervals of approximately 50 units (a uniform distribution), with the exact presented values jittered slightly (e.g., ‘47′ and ‘51′ were presented rather than two instances of ‘50′). Target values used were: 47, 51, 98, 102, 147, 153, 199, 202, 249, 252, 298, 302, 349, 351, 398, 403, 449, 453, 499, 502, 547, 552, 597, 601, 647, 652, 699, 703, 747, 753, 798, 802, 848, 853, 899, 901, 949, and 953. Participants were seated in front of a computer with blank paper covering the keyboard and top of the screen to obscure potential landmarks. Before each task, participants were given written instructions and two practice trials. For each trial, the task was to move the cursor and click the appropriate position (the position that matched the location of the number) on the blank horizontal line during the response screen. Mouse clicks were recorded as numbers from 0 to 1000, corresponding to locations along the response line. A 1000 ms pause separated trials. Each participant completed two blocks (with 38 trials per block × 2 blocks = 76 trials) of the number line estimation task.

2.3.2. Number comparison task

This task (adapted from Dehaene et al., 1990) assesses speed in accurately discriminating magnitudes in the form of numbers. The stimuli consisted of two-digit numerals (3 cm in height), presented in the center of a screen, using Psyscope software. Each numeral was displayed briefly (500 ms) followed by a blank screen (1750 ms). All values from 31 to 99, except 65, were presented once, and then the set was repeated. Two pseudorandom orders of the 138 blocked trials were used. Participants were seated in front of a computer. Before the task, they were given written instructions and five practice trials. For each trial, the task was to press the right-hand response key (‘K’, with right index finger) if the number appearing on the screen was larger than 65, and the left-hand key (‘F’, with left index finger) if the number was smaller than 65, with instructions to respond as quickly and accurately as possible. Responses were collected during the full 2250 ms trial time. Response time and accuracy were recorded.

2.4. Numeracy-related measures

2.4.1. Numeracy scale

The numeracy scale (often called the expanded numeracy scale; Lipkus et al., 2001) is an 11-item measure of one’s ability to manipulate and transform numbers across part-whole formats (e.g., fractions). For example, one question is: “If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?” The scale was administered on paper. This is a commonly used numeracy scale but it has the limitation that scores are typically negatively skewed for college educated samples. We included this scale because in Patalano et al. (2015) numeracy score (log transformed to reduce skew) was correlated with probability weighting curvature, and because the scale specifically assesses part-whole understanding.

2.4.2. Cognitive reflection test

The cognitive reflection test (Frederick, 2005) is a 3-item measure originally designed to capture one’s ability or disposition to resist impulsive, intuitive responding, but has since been identified as a measure of numeracy as well (because task success requires both reflection and number skills; e.g., Campitelli & Gerrans, 2014; Szaszi, Szollosi, Palfi, & Aczel, 2017), and as a predictor of numeracy-related decision making skills (Sinayev & Peters, 2015). An example test question is: “A bat and a ball cost $1.10 in total. The bat costs more than the ball. How much does the ball cost?” The scale was administered on paper. Scores are typically normally distributed for college educated samples. We included this scale because it and the Lipkus et al. (2001) scale taken together are very similar to the 8-item numeracy scale (Weller et al., 2013) used by Schley and Peters (2014) that was found to be correlated with gambling value curvature. That is, all of the items on the Weller et al. (2013) scale except one were drawn from the two scales used here. In the present work, the numeracy scale, cognitive reflection test, and SAT scores will be referred to as numeracy-related measures.

2.5. Gambling task

This task, based on Gonzalez and Wu (1999; see also Patalano et al., 2015; Tversky & Kahneman, 1992), is frequently used to estimate value and probability function parameters in decision making under risk (see Fox & Poldrack, 2014). The stimuli consisted of

(a)

Gamble: 95% change of $100, otherwise $0		
Sure Thing	Prefer Sure Thing	Prefer Gamble
$100	x	
$80	x	
$60		x
$40		x
$20		x
$0		x

Next

(b)

Gamble: 95% change of $100, otherwise $0		
Sure Thing	Prefer Sure Thing	Prefer Gamble
$80	x	
$76	x	
$72	x	
$68		x
$64		x
$60		x

Done

Fig. 3. In the gambling task, participants were instructed to choose a preference between a gamble and each sure-thing option presented. Each trial consisted of: (a) an initial display (with example responses shown as Xs here) and (b) a follow-up display with sure-thing values determined by responses to the initial display. In this example, the certainty equivalent (*CE*) for this gamble would be 70, because this is the midpoint of the dollars where the participant "crossed over" from preferring a sure thing to preferring a gamble on the second display.

165 unique two-outcome gambles created by crossing 15 pairs of dollar values with 11 probabilities. The pairs of dollar values were 25–0, 50–0, 75–0, 100–0, 150–0, 200–0, 400–0, 800–0, 50–25, 75–50, 100–50, 150–50, 150–100, 200–100, and 200–150. The probabilities were 1, 5, 10, 25, 40, 50, 60, 75, 90, 95, and 99. Eleven of these gambles, one at each probability level, were repeated as a reliability check, resulting in a total of 176 gambles (one per trial). Ten orders were created; the trial orders were random except that no gamble was permitted to appear twice in a row.

Each trial of the gambling task began with a display like that in Fig. 3. The task was to compare the stated gamble to each of six "sure-thing" dollar amounts (down the left side of the display) and to determine a preference between the gamble and each sure thing. The sure-thing amounts ranged from the largest possible gamble outcome to the smallest, with intermediate values at equally spaced intervals. Participants were told that it was expected that they would "cross over" from preferring the sure thing to preferring the gamble at some point for each display. When done with the first display, participants pressed a button to go on to a second display for the same trial, to more precisely specify their crossover point for the gamble.

In this second display, the gamble remained the same but the display was refreshed with a narrower range of sure-thing values, and the task was repeated. The endpoints were the sure-thing values on either side of the crossover point from the previous display; for example, given the hypothetical responses in Fig. 3, the endpoints would be $60 and $40. Six intermediate values were again equally spaced between these points. The certainty equivalent for the trial—the monetary outcome identified by an individual as being as attractive as playing a gamble, thus reflecting the overall value of the gamble to the individual—was defined as the dollar value at the participant's crossover point for the second display. For example, if the crossover occurred between $58 and $54, the certainty equivalent was recorded as $56. When done, the participant pressed a button to go on to the next trial.

3. Results

3.1. Missing or excluded data

All participants completed the number comparison task, the numeracy scale, and the cognitive reflection test. A total of 17 participants did not adequately complete the gambling task: 7 did not come for the second session (or left early), 9 had implausible gambling-task response patterns (specifically, for > 30 trials, they did not show a crossover pattern on the first screen of the trial), 1 had a RMSE for the CPT model fit that was more than 4 SD above the mean, and 1 had a repeated-trial response reliability < 0 (see Patalano et al., 2015, for similar exclusion criteria). One additional participant did not complete the number line estimation task. The number of participants excluded based on performance was the same here as in past work (~12% here and in Patalano et al., 2015); however the total exclusions were higher here due to a larger number of participants not returning for the second session. For most analyses reported here, $N = 73$ participants who sufficiently completed all tasks. (The 17 participants excluded for having insufficient data had lower numeracy scores than included participants (9.05 vs. 10.07 respectively, $t(89) = 3.04, p < .003$), however the findings did not change with inclusion of these participants in analyses involving tasks they completed.) The only analyses reported here in which fewer than 73 participants were used were pairwise correlations involving SAT-Math scores. Specifically, a subset of $N = 58$ for whom actual or estimated SAT-Math scores were available were included in SAT-related analyses.

3.2. Measures used with each task or scale

See Table 1 for descriptive statistics for central measures, including number line estimation bias, number comparison slope, numeracy score, cognitive reflection score, SAT-Math score and the three gambling deviation measures. Fits of the models from which

Table 1
Descriptive statistics.

	Mdn	M	SD	Est < 1[1]	Range	Skewness
Scale measures						
Numeracy score	10.00	10.07	1.23	–	5–11	−1.80
Cognitive reflection score	2.00	1.49	1.06	–	0–3	−0.09
Number line estimation task (77% 1-cycle vs. 23% 2-cycle)						
Bias β_d	0.13	0.14	0.07	26%	0.01–0.33	0.49
Model fit R^2	0.98	0.98	0.01	–	0.93–0.99	−1.64
Number comparison task						
Slope b	24.47	27.62	17.39	1%[2]	8–72	0.45
Model fit R^2	0.19	0.21	0.14	–	0–0.53	0.39
Gambling task						
Value curvature α_d	0.32	0.35	0.22	63%	0.01–0.91	0.46
Probability weighting elevation δ_d	0.39	0.46	0.38	77%	0.01–2.21	1.49
Probability weighting curvature β_d	0.54	0.48	0.26	81%	0.05–2.70	0.47
Model fit R^2	0.96	0.95	0.05	–	0.71–1.00	2.90
SAT component score						
SAT-Math	690	685	64	–	490–800	0.55

$N = 73$, except $N = 58$ for SAT measure. [1]Est < 1 refers to the percentage of individuals with an original parameter estimate (not a deviation) < 1; for all β's, < 1 is an inverse S-shaped curve, while > 1 is an S-shaped curve. [2]For b only, what is reported is whether the parameter estimate is less than 0 rather than 1. Prior to further analyses, numeracy score and all deviation measures were log transformed to reduce skew (as in Patalano et al., 2015).

parameter estimates were obtained are also included in Table 1. Prior to further analyses, the numeracy score, number line estimation bias, and the three gambling deviation measures were log transformed to reduce skew for use with parametric tests.

3.2.1. Number line estimation task

An individual's estimate for a target value was removed as an outlier if it differed from the group mean for that value by > 2 SDs (3.8% of trials; per Lai, Zax, & Barth, 2018). Percent absolute error (PAE = |target value – estimate|/response range)*100 was calculated as a general measure of accuracy.[1] The median percent absolute error was 4% (range = 2−7%). Estimation data were then fit with one- and two-cycle versions of the cyclical power model (Hollands & Dyre, 2000; Slusser et al., 2013; Spence, 1990) using Graphpad Prism software. The one-cycle equation is: $y = (x - LB)^{\beta}/((x - LB)^{\beta} + (UB - x)^{\beta})$ where x is the target value and y is the predicted response, LB and UB stand for lower and upper bound respectively, and $LB = 0$ and $UB = 1000$ for all trials. The two-cycle equation is: $y = (x - LB)^{\beta}/((x - LB)^{\beta} + (UB - x)^{\beta}) \cdot 0.5 + LB$, where, for targets between 0 and 500, $LB = 0$ and $UB = 500$, while for targets between 500 and 1000, $LB = 500$ and $UB = 1000$. For each task, we determined whether the one-cycle or two-cycle model better fit a participant's data by selecting the one with the lower AICc score (as in Barth et al., 2015). For model fitting, β was constrained to be between 0 and 2 with a starting value of 0.5. β and model fit R^2 were determined using the best fitting model for each individual. Bias was then computed as a measure of the deviation of curvature from the identity line, independent of direction. This was computed by taking the inverse of all β values greater than 1 (because the curvature associated with an exponent above 1 matches that of its inverse below 1, e.g., $\beta = 2$ and $\beta = ½$ produce the same magnitude of curvature), and then subtracting all values from 1 so that a larger β_d would indicate greater deviation from the identity line. As shown in Table 1, the majority of participants had an S-shaped (rather than inverse S-shaped) curve, and the majority were best fit by a one-cycle instead of a two-cycle model (similar to Barth et al., 2015; Cohen & Blanc-Goldhammer, 2011; Slusser & Barth, 2017).

3.2.2. Number comparison task

The mean number of trials answered correctly was 121 out of 136 (or 88%; $SD = 11$, range = 63–135, however only one score was below 92). Of the trials not answered correctly, a mean of 80% were answered incorrectly and 20% were not answered. Following past procedure (e.g., Dehaene et al., 1990), only response times for correct responses were used in analyses. The response times for the four trials at each distance from 65 were averaged together, where distance = |65-stimulus value|, to obtain a mean response time for each individual for each distance from 65 (a total of 34 distances). The mean of these times was 513 ms ($SD = 72$, range = 418–727; skewness = 0.94). For each individual, the slope of the best fitting line for predicting response time (RT) from distance was computed, using RT = a + -b * ln (distance), where a is the y-intercept, b is the slope, and 'ln' is the natural log. The slope is a measure of discrimination difficulty. To generally compare findings to Dehaene et al. (1990), the best-fitting line using response times averaged over participants was also computed. The fit of the line ($R^2 = 0.81$) was similar to $R^2 = 0.83$ in Dehaene et al. (1990).

[1] Recent work has revealed elevated number line estimation error for target numbers just below 100s boundaries (e.g., 898) compared to those just above (e.g., 901); Lai et al., 2018. We re-ran all analyzes with these trials excluded but findings did not change.

3.2.3. Numeracy-related measures

The numeracy score was computed as the number of correct responses (out of 11), and the cognitive reflection score was computed as the number of correct responses (out of 3). For the participants with standardized test scores available, we either obtained SAT-Math as well as on Critical Reading and Writing (based on pre-2016 SAT format) or (for 6 participants who did not have SAT scores) obtained and converted ACT component scores to SAT equivalents using percentiles. While we do not discuss the verbal component SAT scores in this report, we do note that descriptive statistics were similar for all SAT components, and that the verbal scores were not correlated with any measures here other than SAT-Math. Numeracy scale scores were positively skewed, as anticipated (and thus log transformed), but the cognitive reflection and SAT-Math scores were normally distributed in this sample.

3.2.4. Gambling task

3.2.4.1. Gambling task coding. As described earlier, the certainty equivalent was the midpoint between the dollar values that served as the boundaries of the crossover from choosing a sure thing to choosing a gamble. When there was not a clear crossover (e.g., the individual did not respond or had multiple crossovers), these trials were not used except for ones in which a participant indicated a crossover point on the first screen of a trial. In these cases, because the certainty equivalent range was narrowed by the first response, it was reasonable to conclude that a participant who chose all sure-thing responses on the second screen was intending the lowest crossover point while a participant who chose all gamble responses intended the highest (following Patalano et al., 2015). Using this procedure, an average of 2.4 trials ($SD = 4.59$, range = 0–24) per participant were excluded. The mean reliability across repeated trials was $r = 0.95$ ($SD = 0.10$, range = 0.48–1.00; only 3 scores were < 0.70), indicating consistency of response and suggesting that attention was paid to the task.

3.2.4.2. Parameter estimation procedure. The following general cumulative prospect theory (CPT) equation was used to model behavior (Tversky & Kahneman, 1992): $v(CE) = v(x_1)w(p) + v(x_2)(1 - w(p))$. In this equation, $v(x)$ represents the value function that transforms the certainty equivalent (CE) and each dollar value (x_1 and x_2) into subjective values, and $w(p)$ represents the probability weighting function that transforms the probability p associated with x_1 (the larger dollar value) into a decision weight (the decision weight for the second probability is 1 minus the weight of the first probability). For the value function, $v(x) = x^{\alpha}$ was used (Tversky & Kahneman, 1992) and, for the probability weighting function, $w(p) = \delta^{\beta}/(\delta p^{\beta} + (1 - p)^{\beta})$ was used (Lattimore, Baker, & Witte, 1992; see also Gonzalez & Wu, 1999). The resulting CPT equation used here for non-linear regression was: $CE = (x_1{}^{\alpha} \cdot \delta p^{\beta}/(\delta p^{\beta} + (1 - p)^{\beta}) + x_2{}^{\alpha} \cdot (1 - \delta p^{\beta}/(\delta p^{\beta} + (1 - p)^{\beta})))^{1/\alpha}$. In this equation, α indexes value curvature, β probability weighting curvature, and δ probability weighting elevation. To obtain parameter estimates, we fit the equation to each participant's CE data for the 165 unique trials. Nonlinear least squares regression and a sequential quadratic programming algorithm (with SPSS statistical software) were used. Estimates were constrained to > 0.001 with starting values of 0.8.[2]

3.2.4.3. Parameter estimates and deviations. Descriptive statistics for parameter estimates are in Supplemental Materials; medians are similar to those reported in past work (e.g., Patalano et al., 2015). The median percentage of variance explained by the cumulative prospect theory model was 96% relative to 70% explained by an expected value model (i.e., a model in which all parameters are set to 1), indicating good fit of the cumulative prospect theory model. As shown in Table 1, the majority (but not all) of participants had a concave curve for the value function, and an inverse-S shaped curve for the probability weighting function, with a crossover of the inverse-S shaped curve below the middle of the probability scale (i.e., below 50%). Using the parameter estimates, we created deviations. For the two parameters that are exponents (α and β), we took the inverse of all estimates above 1, and then subtracted all values from 1 so that a larger α_d and β_d would indicate greater deviation from the identity line (i.e., greater curvature or bias) without regard to direction. For δ, which is a multiplier rather than an exponent, we took the absolute value of the parameter estimate minus 1 (that is, $\delta_d = |1 - \delta|$) to get the distance of the curve from the identity line midpoint (i.e., away from a crossover of 50%) without regard to direction (that is, whether the curve was raised or lowered).

3.3. Correlational analyses

3.3.1. Correlations between predictor measures

As shown in Table 2, number line estimation bias was not correlated with number comparison slope, and neither measure was reliably correlated with numeracy, cognitive reflection, or SAT-Math scores. Numeracy-related measures were moderately positively correlated with one another, specifically, numeracy and cognitive reflection were correlated ($r(56) = 0.31$, $p = .006$), as were cognitive reflection and SAT-Math ($r(56) = 0.56$, $p < .001$). Because gender differences in number skills are generally of interest, we report them in correlational analyses. Notably, men had higher scores than women on cognitive reflection ($r(71) = 0.28$, $p = .018$) and SAT-Math ($r(56) = 0.30$, $p = .023$).

[2] While we previously developed a version of the probability weighting function that accommodates two-cycle curves (see Xing et al., 2019), similar to the two-cycle number line estimation curve, we did not use it here given little evidence to date that it offers a better fit to individual-level gambling data, and because identifying which model is preferred is not straightforward in the context of a very flexible (with many parameters) CPT equation.

Table 2
Pearson correlations for number skills measures (and gender).

	GEND	NUM	CRT	NL β_d	NC b
Gender	–				
Numeracy score	0.17	–			
Cognitive reflection score	**0.28***	**0.31****	–		
Number line estimation bias β_d	0.16	-0.02	0.04	–	
Number comparison slope b	-0.04	-0.01	-0.06	0.02	–
SAT-Math	**0.28***	0.12	**0.55*****	-0.13	0.20

$N = 73$, except $N = 58$ for SAT-Math. *Notes:* Gender identity is coded as 1 = woman and 2 = man.
*** $p < .001$.
** $p < .01$.
* $p < .05$.

Table 3
Pearson correlations with gambling deviation measures.

	Gambling deviation measures		
	Value curvature α_d	Probability weighting elevation δ_d	Probability weighting curvature β_d
Gender	0.05	0.09	−0.03
Numeracy score	**−0.29***	−0.23	−0.19
Cognitive reflection score	−0.06	−0.14	**−0.24***
Number line estimation bias β_d	0.07	0.03	**0.32****
Number comparison slope b	0.18	**0.39****	−0.18
SAT-Math	−0.04	0.09	**−0.37****

$N = 73$, except $N = 58$ for SAT-Math.
** $p < .01$.
* $p < .05$.

3.3.2. Correlations between predictors and gambling deviations

As shown in Table 3, there were reliable correlations between gambling deviations and number tasks. As predicted, individuals with greater probability weighting curvature also had greater number line estimation bias ($r(71) = 0.32$, $p = .007$) but did not have greater number comparison slope ($r(71) = -0.18$, $p = .134$). In contrast, individuals with greater probability weighting elevation did not have greater number line estimation bias, $r(71) = 0.03$, $p = .825$ as predicted, but surprisingly did have greater number comparison slope, $r_s(71) = 0.39$, $p = .001$. Gambling value curvature was not reliably correlated with either number line estimation bias ($r(71) = 0.07$, $p = .567$) or number comparison slope ($r(71) = 0.18$, $p = .127$). The findings are consistent with probability weighting curvature being related to proportion judgment in that degree of curvature was predicted only by number line estimation bias.[3]

As expected, there were also reliable correlations between gambling deviations and numeracy-related measures. As shown in Table 3, individuals higher in cognitive reflection and SAT-Math had less probability weighting curvature, while individuals higher in numeracy had less value curvature. None of the numeracy-related measures predicted probability elevation. We speculate that the differences in correlations involving the numeracy-related measures might be in part due to test difficulty. That is, numeracy might have better differentiated performance of low-numeracy participants from others across a wide range of tasks, while cognitive reflection and SAT-Math might have better differentiated performance of more highly numerate individuals on more complex tasks. Because SAT scores were not required for college admission, it is likely that the subsample submitting SAT-Math scores had stronger math skills (and thus that the set of scores was less variable) than the full sample. It is also possible that the measures tap into different skills, and that some skills are more relevant to interpretation of value-related or probability-related information during decision making.

3.3.3. Correlations regarding direction of curvature

Because we found evidence of a relationship between number line estimation bias and probability weighting curvature, we also tested whether there was a relationship in the direction of each curve. We found that whether the number line parameter estimate was S-shaped or inverse S-shaped did not predict the direction of the gambling probability weighting curve, $\chi^2(1, N = 73) = 1.19$, $p = .275$). Because we found a relationship between number comparison slope and probability elevation, we also tested whether slope predicted the direction of the elevation. Again, it did not; individuals with greater slope were not more likely to have a raised

[3] We also re-did gambling parameter estimation using the one-parameter probability weighting function used by Schley and Peters (2014), namely, $w(p) = \exp[-(-\ln(p))^{\beta}$ (developed by Prelec, 1998), and re-ran correlations between gambling deviation scores and number line estimation bias. The pattern of findings was unchanged.

Table 4
Linear regressions for predicting gambling deviation measures.

	b	*SE*	*β*	*t*	*p*
Value curvature α_d					
Constant	0.07	0.03		2.67	.010
Numeracy score	**−0.09**	**0.04**	**−0.30**	**−2.53**	**.014**
Cognition reflection score	0.01	0.01	−0.04	0.37	.713
Number line estimation bias β_d	0.14	0.29	0.06	0.49	.626
Number comparison slope *b*	0.01	0.01	0.18	1.56	.124
Probability weighting elevation δ_d					
Constant	0.07	0.04		1.94	.056
Numeracy score	−0.09	0.05	−0.20	−1.79	.077
Cognition reflection score	−0.01	0.01	−0.05	−0.44	.661
Number line estimation bias β_d	0.05	0.42	0.12	0.11	.913
Number comparison slope *b*	**0.01**	**0.01**	**0.38**	**3.49**	**.001**
Probability weighting curvature β_d					
Constant	0.15	0.03		5.20	< .001
Numeracy score	−0.04	0.04	−0.11	−1.02	.310
Cognition reflection score	**−0.02**	**0.01**	**−0.23**	**−2.10**	**.048**
Number line estimation bias β_d	**0.98**	**0.32**	**0.33**	**3.06**	**.003**
Number comparison slope *b*	−0.01	0.01	−0.20	−1.90	.062

$N = 73$. *Notes:* Model fit for gambling α_d was $F(4,72) = 2.35, p = .063, R^2 = 0.12$; for δ_d was $F(4,72) = 4.24, p = .004, R^2 = 0.20$; and for β_d was $F(4, 72) = 4.69, p = .002, R^2 = 0.22$. Statistically significant predictors are bolded above. Notably, gambling β_d was predicted by number line estimation bias β_d, while gambling δ_d was predicted by number comparison slope *b*.

curve (with a crossover above 50%) than a lowered curve (i.e., they were not more inclined towards risk; $r(71) = -0.06, p = .588$). We also considered whether the complexity of the best fitting number line cyclical power model (one vs. two cycles) predicted probability weighting curvature (i.e., β_d) but it did not, $r(71) = 0.07, p = .543$. The findings to this point are consistent with the central prediction of a relationship in degree of bias across proportion judgment tasks.

3.4. Linear regression for predicting gambling deviations

In a follow-up analysis (see Table 4), all number-related predictors for which we had data from all participants (i.e., not gender because it was not a number-related predictor, and not SAT-Math because we did not have scores for all participants) were entered into linear regressions for predicting the gambling deviations. We ran these analyses for two reasons. First, we wanted to confirm that number line estimation bias and number comparison slope remained strong predictors even in the context of broader numeracy-related measures. Second, we wanted to assess whether broader numeracy-related measures would still be predictive in the context of the more specific number skills tasks. If not, the relationship between gambling deviations and numeracy-related measures could be said to be explained by the specific number skills.

For probability weighting curvature, number line estimation bias remained a strong positive predictor of deviation ($p = .003$) and cognitive reflection remained in the model (that is, that the relationship was not fully explained by number line estimation skills; $p = .048$), consistent with the correlational analysis. For probability weighting elevation, only number comparison slope was a strong positive predictor ($p = .001$) of deviation, again consistent with the correlational analyses. Finally, for value curvature, as with the correlational analyses, only numeracy was a statistically reliable predictor ($p = .004$). These findings provide further evidence that it is the specific number tasks rather than broader numeracy that drive the task-related correlations with gambling deviations seen here (but that there also remain numeracy-related correlations not explained by performance on the number skills tasks).

4. Discussion

The present study was motivated by the observation that, like number line estimation, estimation of probabilities expressed as symbols can be conceptualized as proportion judgment, and that there might be a common psychological explanation for the S-shaped and inverse S-shaped curves seen across these and a wide range of other proportion judgment tasks. We found evidence of a double dissociation: number line estimation bias was correlated with probability weighting curvature in a gambling task (when both were modeled using similar proportion judgment equations), while number comparison (a task that does not involve proportion judgment) was correlated with probability weighting elevation. Specifically, the more estimation bias there was in the number line task, the more gambling probability curvature there was as well (although there was no relationship across tasks in the direction of the curve). This finding is consistent with proportion judgment underlying the canonical inverse S-shaped (or sometimes S-shaped) pattern of probability distortion in decision making. That is, the probability weighting curve's shape (and individual differences in shape) might arise, at least in part, from a part-whole estimation involving two inexact representations of magnitude (e.g., 75 out of 100). This finding is not only evidence that symbolic number skills are related to probability weighting curvature, but it also suggests a potential explanation for the shape of the probability weighting curve.

This central finding is supported by recent work of Müller-Trede et al. (2018) who found that it is the bounded nature of the probability scale (and not its substantive content) that is critical to the shape of the curve. The proportion judgment approach is also in the spirit of Tversky and Kahneman (1992) psychophysical explanation for the probability weighting curve, except that while proportion judgment focuses on the relationship between two imprecise magnitudes (following Hollands & Dyre, 2000; Spence, 1990), Tversky and Kahneman (1992) focused on diminishing sensitivity from two reference points (i.e., 0 or 1). The latter does not easily accommodate the regular (not-inverse) S-shape probability weighting curve that sometimes arises, and cannot explain more complex patterns in other tasks (e.g., multi-cycle patterns that have been found in number line estimation). Diminishing sensitivity is thus less likely to be able to offer a unifying account. That said, it is difficult to provide evidence to discriminate between accounts because, while they are distinct psychological explanations, the mathematical equations used in modeling are typically similar if not the same. The present study reflects one route to differentiating between accounts.

Interestingly, we did not observe commonality in the direction of curvature across proportion judgment tasks. Because individual differences in the direction of curvature are not well understood, we are not able to fully assess the meaning of this result. However, it has been demonstrated with perceptual stimuli that whether discriminability increases or decreases as stimulus magnitude increases depends on the type of stimulus being judged (e.g., loudness, visual area, distance; see Stevens, 1957; Hollands & Dyre, 2000). It is possible that different tasks in the present study evoked different types of judgments, that is, that symbolic numbers might be mapped to magnitudes differently when the numbers reflect risk assessments. Alternatively, underlying magnitudes might be similarly estimated but tasks might be construed differently, leading to different proportion judgments. For example, this might happen if an individual consistently reframed likelihoods (e.g., reconstruing a 99% chance of getting $800 as a 1% chance of not getting it) and estimated the magnitude of the inverse of given values. We also note that the two proportion judgment tasks used here, while both involving symbolic number stimuli, used different response modalities (i.e., a physical line vs. a choice between pairs of options). The mapping of the stimulus to response format matters in general, and might matter here as well: for example, the direction of bias in adults flips to inverse-S shaped when the task is to map a number line location to a number (the reverse of what was done here; Castronovo, Crollen, & Seron, 2010; Slusser & Barth, 2017).

One question one might ask is if bias across tasks ultimately reflects underlying magnitudes, why is number line estimation bias not also related to gambling value curvature? We suspect that it probably is, given that a reliable correlation was observed in past work (Peters & Bjalkebring, 2015; Schley & Peters, 2014). It may be that the correlation was stronger between number line estimation bias and probability weighting curvature here because these tasks involve both magnitude estimation and proportion judgment skills. In other words, a large bias on either task may reflect an inexactness with both skills, rather than only with magnitude estimation (as suggested by Cohen & Sarnecka, 2014; Schneider et al., 2018). Somewhat relatedly, one might ask why β is typically greater than 1 in both unbounded and bounded number line estimation tasks in adults, which seems surprising given other evidence that discrimination is more difficult as magnitudes increase (suggesting $\beta < 1$). Even if one does not take the unbounded number line task as evidence of underlying magnitude representation, the same $\beta > 1$ is found in other unbounded tasks such as when one must generate a dot pattern that reflects a given numerical magnitude (Krueger, 1984), so it does not appear to be specific to number line estimation. Teasing apart contributors to bias, and modeling magnitude representation, are two important issues for future research.

We interpret the findings of the present study with caution because they differ from those of Schley and Peters (2014) who reported a correlation between number line estimation error (i.e., percent absolute error) and gambling value curvature but not probability weighting (using a different gambling task and a different probability weighting function). The discrepancy cannot be attributed to differences in probability weighting models (as we reanalyzed our data using the model they used) or to number line estimation measures (as they indicated in a footnote also computing a bias measure for number line estimation similar to the one used here). Schley and Peters proposed that predictive power might load on one parameter or the other during modeling, and that this might explain the lack of a correlation with the probability weighting parameter estimate in their own work. We suggest that the loadings might also depend on the procedure used to estimate parameters (the adaptive choice procedure vs. the certainty equivalent procedure), which is quite likely given that different combinations of parameter estimates can well describe the same pattern of behavior (Stott, 2006). The number line estimation task used here—with whole-number targets distributed over the full 0 to 1000 range—might have led to different number line estimation strategies between the two studies.

A surprising present finding was that the number comparison slope was most strongly correlated with probability weighting elevation rather than value curvature or probability weighting curvature. The number comparison and number line estimation tasks differ in that the latter elicits judgments of magnitude on a ratio scale while the former elicits ordinal comparisons. Performance measures for the two tasks are not consistently correlated, even among children, suggesting that they draw on only partially overlapping skills (see Schneider et al., 2017, for a review). It is not surprising that the two measures were differently related to decision making here, but it is unclear why number comparison skills would be specifically related to probability weighting elevation (which has been associated with the overall attractiveness of gambling to the individual; Gonzalez & Wu, 1999). One possible interpretation of the findings is that individuals who have greater slope on the number comparison task are less skilled in linking their overall intuitive assessment of a gamble to a certainty equivalent value and that this is what is captured by the elevation parameter in this context. What is clear is that probability weighting elevation may be as related to number skills as are other gambling measures, and that number skills tasks cannot be treated as interchangeable assessments of relevant skill (see Peters & Bjalkebring, 2015, for a similar point).

We included several numeracy-related scale measures here, and replicated that numeracy-related measures predict biases in gambling (Schley & Peters, 2014; Patalano et al., 2015) and SAT-Math (Frederick, 2005). One perhaps surprising finding is that the numeracy-related measures (including SAT-Math) were not correlated with number skills measures. Moderate correlations between the latter and various math achievement scores are frequently found (rs $= \sim 0.30$; see Schneider et al., 2018; Schneider et al., 2018, for meta-analyses), with effect sizes higher for number line estimation than number comparison. However, the majority of these studies focus on children, which makes comparison difficult. Effect sizes have also been shown to vary greatly as a function of specific number task and mathematical competence measures used. For example, effect sizes involving number comparison slope are typically much smaller than those using other measures, so it is difficult to compare across measures as well. We suspect that we did not replicate past correlations because the sample consisted of highly numerate adults, but it might also be that the measures used here are less strongly related to one another than are other measures used in past work.

In conclusion, there already exists overwhelming existing evidence that numeracy is related to the use of number-based strategies during decision making (see Peters et al., 2006; Reyna et al., 2009, for reviews), and there is growing evidence that performance measures on tasks that tap into symbolic number skills are specifically related to bias in the use of numbers during decision making (Schley & Peters, 2014; Peters et al., 2006; Peters, Hart, Tusler, & Fraenkel, 2014). But there has been less work focused on explaining the shape of the probability weighting function in general, and also on individual differences in its curvature and elevation as a function of intuitive number skills. The present work draws on a proportion judgment approach to motivate the shape of the probability weighting curve (see also Xing, Paul, Zax, Cordes, Barth, & Patalano, 2019), it provides initial evidence of a relationship between the magnitudes of proportion judgment biases across the two tasks used here, and it more generally contributes to evidence of a shared psychological explanation for patterns of behavior across tasks with bounded scales.

5. Author note

We thank lab members Chenmu Xing, Marcia Mata, and Lily Segal for their valuable assistance, and Dr. Emily Slusser for programming one of the tasks. This work was supported by funding from the National Science Foundation DRL-1561214 and Wesleyan University in the United States.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2020.101273.

References

Alempaki, D., Canic, E., Mullett, T. L., Skylark, W. J., Starmer, C., Stewart, N., & Tufano, F. (2019). Reexamining how utility and weighting functions get their shapes: A quasi-adversarial collaboration providing a new interpretation. *Management Science, 65*, 4841–4862.

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology, 111*, 246–267.

Barth, H., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14*, 125–135.

Barth, H., Lesser, E., Taggart, J., & Slusser, E. B. (2015). Spatial estimation: A non-Bayesian alternative. *Developmental Science, 18*, 853–862.

Birnbaum, M. H. (1974). Reply to the devil's advocates: Don't confound model testing and measurement. *Psychological Bulletin, 81*, 854.

Black, W. C., Nease, R. F., Jr., & Tosteson, A. N. (1995). Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *JNCI: Journal of the National Cancer Institute, 87*, 720–731.

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition, 42*, 434–447.

Castronovo, J., Crollen, V., & Seron, X. (2010). Visual enumeration: A bi-directional mapping process between symbolic and non-symbolic representations of number? *Journal of Vision, 10* 1327-1327.

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review, 18*, 331–338.

Cohen, D. J., Blanc-Goldhammer, D., & Quinlan, P. T. (2018). A mathematical model of how people solve most variants of the number-line task. *Cognitive Science, 42*, 2621–2647.

Cohen, D. J., & Sarnecka, B. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology, 50*, 1640–1652.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making, 7*, 25–47.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences, 7*, 145–147.

Dehaene, S. (1997). *The number sense.* New York: Oxford University Press.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1–42.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 626–641.

Fox, C. R., & Poldrack, R. A. (2014). Prospect theory and the brain. In P. Glimcher, & E. Fehr (Eds.). *Neuroeconomics: Decision making and the brain* (pp. 533–567). (2nd ed.). London: Elsevier.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25–42.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43–74.

Gallistel, C. R., & Gelman, R. (2005). *Mathematical cognition.* New York: Cambridge University Press.

Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition, 123*, 21–32.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology, 38*, 129–166.

Hogarth, R. M., & Einhorn, H. J. (1990). Venture theory: A model of decision weights. *Management Science, 36*, 780–803.

Hollands, J. G., & Dyre, B. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review, 107*, 500–524.

Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & Psychophysics, 35*, 536–542.

Lai, M., Zax, A., & Barth, H. (2018). Digit identity influences numerical estimation in children and adults. *Developmental Science, 21*, e12657.

Landy, D., Guay, B., & Marghetis, T. (2018). Bias and ignorance in demographic perception. *Psychonomic Bulletin & Review, 25*, 1606–1618.

Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of Economic Behavior & Organization, 17*, 377–400.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*, 37–44.

Meck, W. H., Church, R. M., & Gibbon, J. (1985). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 11*, 591.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature,* 519–520.

Müller-Trede, J., Sher, S., & McKenzie, C. R. (2018). When payoffs look like probabilities: Separating form and content in risky choice. *Journal of Experimental Psychology: General, 147*, 662–670.

Nakajima, Y. (1987). A model of empty duration perception. *Perception, 16*, 485–520.

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology, 91*, 191–206.

Patalano, A. L., Saltiel, J. R., Machlin, L., & Barth, H. (2015). The role of numeracy and approximate number system acuity in predicting value and probability distortion. *Psychonomic Bulletin & Review, 22*, 1820–1829.

Patalano, A. L., Lolli, S. L., & Sanislow, C. A. (2018). Gratitude intervention modulates P3 amplitude in a temporal discounting task. *International Journal of Psychophysiology. 103*, 202–210.

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

Peters, E., & Bjalkebring, P. (2015). Multiple numeric competencies: When a number is not just a number. *Journal of Personality and Social Psychology, 108*, 802.

Peters, E., Hart, P. S., Tusler, M., & Fraenkel, L. (2014). Numbers matter to informed patient choices: A randomized design across age and numeracy levels. *Medical Decision Making, 34*, 430–442.

Peters, E., Slovic, P., Västfjäll, D., & Mertz, C. K. (2008). Intuitive numbers guide decisions. *Judgment and Decision Making, 3*, 619–635.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science, 17*, 407–413.

Prelec, D. (1998). The probability weighting function. *Econometrica, 60*, 497–528.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin, 135*, 943–973.

Rottenstreich, Y., & Hsee, C. S. (2001). Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological Science, 12*, 185–190.

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science, 17*, 525–536.

Schley, D. R., & Peters, E. (2014). Assessing economic value: Symbolic-number mappings predict risky and riskless valuations. *Psychological Science, 25*, 753–761.

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2018a). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science, 20*, e12372.

Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018b). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development, 89*, 1467–1484.

Schneider, M., Thompson, C. A., & Rittle-Johnson, B. (2017). Associations of magnitude comparison and number line estimation with mathematical competence: A comparative review. In P. Lemaire (Ed.). *Cognitive development from a strategy perspective: A Festschrift for Robert S.Siegler* (pp. 100–119). London: Routledge.

Shuford, E. H. (1961). Percentage estimation of proportion as a function of element type, exposure time, and task. *Journal of Experimental Psychology, 61*, 430.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237–250.

Simpson, W., & Voss, J. F. (1961). Psychophysical judgments of probabilistic stimulus sequencies. *Journal of Experimental Psychology, 62*, 416.

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology, 6*, 532.

Slusser, E. B., & Barth, H. (2017). Intuitive proportion judgment in number-line estimation: Converging evidence from multiple tasks. *Journal of Experimental Child Psychology, 162*, 181–198.

Slusser, E. B., Santiago, R., & Barth, H. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General, 142*, 193–208.

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

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153–181.

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *Quarterly Journal of Experimental Psychology, 62*, 1041–1062.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology, 53*, 1–26.

Stewart, N., Reimers, S., & Harris, A. J. L. (2015). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science, 61*, 687–705.

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty, 32*, 101–130.

Sullivan, J., Juhasz, B., Slattery, T., & Barth, H. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review, 18*, 557–563.

Szaszi, B., Szollosi, A., Palfi, B., & Aczel, B. (2017). The cognitive reflection test revisited: Exploring the ways individuals solve the test. *Thinking & Reasoning, 23*, 207–234.

Toubia, O., Johnson, E., Evgeniou, T., & Delquié, P. (2013). Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. *Management Science, 59*, 613–640.

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

Varey, C. A., Mellers, B. A., & Birnbaum, M. H. (1990). Judgments of proportions. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 613.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making, 26*, 198–212.

Winman, A., Juslin, P., Lindskog, M., Nilsson, H., & Kerimi, N. (2014). The role of ANS acuity and numeracy for the calibration and the coherence of subjective probability judgments. *Frontiers in Psychology, 5*, 851.

Xing, C., Paul, J., Zax, A., Cordes, S., Barth, H., & Patalano, A. L. (2019). Probability range and probability distortion in a gambling task. *Acta Psychologica, 197*, 39–51.

Zeisberger, S., Vrecko, D., & Langer, T. (2012). Measuring the time stability of prospect theory preferences. *Theory and Decision, 72*, 359–386.

Zhang, H., & Maloney, L. T. (2012). Ubiquitous log odds: A common representation of probability and frequency distortion in perception, action, and cognition. *Frontiers in Neuroscience, 6*, 1.