# The causal impact of objective numeracy on judgments: Improving numeracy via symbolic and non-symbolic approximate arithmetic training yields more consistent risk judgments

Dana Chesney[a], Brittany Shoots-Reinhard[b] & Ellen Peters[c]

[a] Department of Psychology, St. John's University, 8000 Utopia Pkwy, Jamaica, NY 11439, USA.

[b] Department of Psychology, The Ohio State University, 1827 Neil Ave., Columbus, OH 43210, USA.

[c] School of Journalism and Communication, Allen Hall, 1275 University of Oregon, Eugene, OR 97403, USA.

Manuscript resubmitted to *Journal of Numerical Cognition,* May 21, 2021

## Author Note

Correspondence may be addressed to: Dana Chesney at St. John's University, Dept. of Psychology, 8000 Utopia Pkwy, Jamaica, NY 11439, USA. email: chesneyd@stjohns.edu, Tel.: 1-718-990-5056. Dana Chesney and Ellen Peters were at The Ohio State University while this research was conducted. The research was supported by NSF grant SES-1155924. We thank Martin Tusler for his assistance. We thank William Chaplin for his help with Bayesian analyses. We thank Melissa Peckins for her help with power analyses. Note: Some of these findings were presented at the 2016 annual conference for the Society for Judgment and Decision Making.

**Abstract**

Park and Brannon (2013, 2014) found that practicing non-symbolic approximate arithmetic increased performance on an objective numeracy task, specifically symbolic arithmetic. Manipulating objective numeracy would be useful for many researchers, particularly those who wish to investigate causal effects of objective numeracy on performance. Objective numeracy has been linked to performance in multiple areas, such as judgment-and-decision-making (JDM) competence, but most existing studies are correlational. Here, we expanded upon Park and Brannon's method to experimentally manipulate objective numeracy and we investigated whether numeracy's link with JDM performance was causal. Experimental participants drawn from a diverse internet sample trained on approximate-arithmetic tasks whereas active control participants trained on a spatial working-memory task. Numeracy training followed a 2×2 design: Experimental participants quickly estimated the sum of OR difference between presented numeric stimuli, using symbolic numbers (i.e., Arabic numbers) OR non-symbolic numeric stimuli (i.e., dot arrays). We partially replicated Park and Brannon's findings: The numeracy training improved objective-numeracy performance more than control training, but this improvement was evidenced by performance on the Objective Numeracy Scale, not the symbolic arithmetic task. Subsequently, we found that experimental participants also perceived risks more consistently than active control participants, and this risk-consistency benefit was mediated by objective numeracy. These results provide the first known experimental evidence of a causal link between objective numeracy and the consistency of risk judgments.

*Key words: objective numeracy, decision making competence, risk perception, causal effects, subjective numeracy, symbolic number mapping, approximate number sense*

# The causal impact of objective numeracy on judgments: Improving numeracy via symbolic and non-symbolic approximate arithmetic training yields more consistent risk judgments

People's numerical ability (i.e., numeracy) has been shown to predict outcomes in a wide variety of areas including academic achievement, income, and health (Adelman, 2006; Bynner & Parson, 2009; National Council of Teachers of Mathematics, 2000; National Mathematics Advisory Panel, 2008; National Research Council, 2001; Peters, Meilleur, & Tompkins, 2013; Peters, Tompkins, Knoll, Ardoin, Shoots-Reinhard, & Meara, 2019; Reyna, Nelson, Han, & Dieckmann, 2009; Smith, McArdle, & Willis, 2010). More numerate people make more – and often better – use of numbers when making judgments (Chesney & Obrecht, 2012; Obrecht & Chesney, 2013; Peters, 2012; Peters & Levin, 2008; Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2006). For example, more numerate individuals are less susceptible to framing effects and show greater sensitivity both to expected value and to different levels of probability (Barton, Cokely, Galesic, Koehler, & Haas, 2009; Obrecht & Chesney, 2013; Peters, 2012; Peters et al., 2006; Reyna, et al., 2009). Intuitively, it makes sense that this link would be causal: Judgments and decisions often involve numbers, so it should follow that those who are better with numbers also will make more use of these numbers in judgment and decision tasks. However, to date, research on this link has been correlational (with one exception: Peters et al., 2017), with investigators examining decision making among individuals who vary in numeracy. The present study begins to address this gap in the literature, using a numeracy intervention recently developed by Park and Brannon (2013, 2014) to manipulate individuals' numeric abilities and examine subsequent effects.

**Manipulating Numeracy**

If more numerate individuals make more use of numbers when making decisions *because* they are more numerate, then increasing individuals' numeracy should yield more number use in decisions. However, increasing numeracy is typically neither quick nor easy. The traditional method of manipulating numeracy – education – typically involves frequent instructor/student interaction over a period of months, if not years. This approach is infeasible in standard research contexts and is not conducive to random assignment of participants from the general population. (One example of this approach, however, can be seen in Peters et al. (2017), a 9-week longitudinal study following students taking a statistics course required for undergraduate psychology majors. They found that a manipulation – designed to decrease threat responses in the context of a statistics course – increased objective numeracy scores as predicted and had consequences for health and financial outcomes.) Moreover, interventions focused on specific skills (which are often more practical to implement) show limited evidence of transfer to other domains (e.g., Chesney & McNeil, 2014). Even transfer between arithmetic problems can be difficult to achieve (Perry, 1991).

Recently, Park and Brannon (2013, 2014) developed an intervention that improved participants' numeracy within a relatively small number of practice sessions. Experimental participants practiced non-symbolic approximate arithmetic, doing mental addition and subtraction on sets of dots. The task was dynamic, adapting to become more or less difficult as participants answered correctly or incorrectly. These experimental participants showed more improvement on an exact symbolic arithmetic task than control participants who completed a spatial-memory training task.

This finding is important for multiple reasons. First, the method is a relatively easy-to-

implement way to manipulate numeracy and, thus, is beneficial for researchers seeking to test causal effects of numeracy on various outcomes (e.g., judgment and decision making, JDM, performance). Second, the inexact, non-symbolic arithmetic practice transferred to a *prima facie* quite different exact symbolic arithmetic task. Park and Brannon's (2013, 2014) training task therefore might be more likely to yield cross-domain benefits than more traditional training. Third, Park and Brannon's method seems to invoke multiple aspects of numeracy as reviewed next.

**Multiple Aspects of Numeracy**

Numeracy is not a single construct (Peters & Bjälkebring, 2015). Instead, it is composed of several inter-related but separable components that have sometimes been used interchangeably in the literature. These include *objective numeracy* (the ability to use and understand mathematical concepts, such as probability and arithmetic: Weller, Dieckmann, Tusler, Mertz, Burns, & Peters, 2013), *subjective numeracy* (beliefs about and attitudes toward numbers, including self-evaluation of numeric ability: Fagerlin, Zikmund-Fisher, Ubel, Jankovic, Derry, & Smith, 2007), *Approximate Number System (ANS) acuity* (an inexact ability to perceive numeric magnitudes from non-symbolic sets, such as dots, without counting; Kaufman, Lord, Reese, & Volkmann, 1949; Taves, 1941), and *symbolic number mapping* (SMap, the ability to map symbolic numbers to numerical magnitudes and understand their relative magnitudes; Chesney, Bjälkebring, & Peters, 2015; Peters & Bjälkebring, 2015; Siegler & Opfer, 2003).

These numeracy components are linked. The ANS, for example, provides a "feel" for the quantities referred to by symbolic numbers, such that symbolic number mapping reflects ANS-acuity and the connection between symbolic numbers and ANS magnitudes (Chesney et al., 2015; Dehaene, 1992). In other words, we learn to map the number "20" to the magnitude the

ANS perceives from a set of 20 items. Mapping precision is thus informed by the precision of the ANS, with greater ANS-acuity being related to better mapping precision (Chesney et al., 2015; Schley & Peters, 2015). People with better ANS-acuity and mapping precision tend to be both more objectively and subjectively numerate (Chen & Li, 2014; Chesney et al., 2015; Peters & Bjälkebring, 2015).

Park and Brannon's (2013, 2014) results demonstrate that training on non-symbolic approximate arithmetic – a task that heavily relies on the ANS – transferred successfully to an objective numeracy task consisting of standard arithmetic problems with symbolic numbers. Moreover, the improvement was not due to practicing specific symbolic arithmetic problems since no such problems were practiced. Rather, the benefit appeared to result from a broader numeric-ability improvement. Of importance here, successful manipulation of objective numeracy would allow us to experimentally investigate the hypothesized causal link between numeric ability and decision performance.

**The Current Study**

In the present study, we used a version of Park and Brannon's (2013, 2014) training task to experimentally manipulate objective numeracy. We modified the original task in three ways. First, we ran our study online using a diverse internet sample, rather than training a small university-based sample in the lab. This more diverse sample allowed us to see if the training benefit would scale to a broader population. Second, participants in a numeracy-training condition completed either addition OR subtraction estimations rather than interleaving trials as in the original studies (piloting demonstrated the non-symbolic addition and subtraction tasks had different difficulty levels, preventing appropriate calibration of the dynamic-training difficulty when run together). Third, we added additional training groups where participants

practiced with Arabic numbers instead of dot sets. This change allowed us to explore differences in symbolic vs. non-symbolic training effects. In particular, non-symbolic training may be more helpful to individuals who lack confidence in their math abilities (i.e., those who are lower in subjective numeracy; Fagerlin et al., 2007) because lack of confidence tends to undercut performance. In fact, researchers have found that people who are less subjectively numerate are more avoidant of math courses and math content, learn less math, and have poorer comprehension on everyday decision-related tasks than those who are higher in subjective numeracy (Ashcraft, 2002; Betz, 1978; Läg, Bauger, Lindberg, & Friborg, 2014; Maloney & Beilock, 2012; Rolison, Morsanyi, & O'Connor, 2016). This effect may be alleviated, however, by reducing the "threat" of math contexts and/or supporting numeric self-efficacy (Maloney & Beilock, 2012; Peters et al., 2017). Non-symbolic training may offer an alternative pathway through which less math-confident individuals can improve their objective numeracy, while avoiding the "threat" and decreased math self-efficacy and persistence that are invoked by traditional math tasks (Betz & Hackett, 1981; Hackett & Betz, 1981).

Like Park and Brannon (2013, 2014), we assessed objective numeracy at pretest and posttest. We expected to replicate their finding that estimation practice improves objective numeracy. However, we also examined possible effects of our intervention on subjective numeracy, symbolic number mapping, and JDM performance. This approach allowed us to determine whether the benefit of the intervention was specific to objective numeracy and to address additional novel hypotheses.

**Hypotheses**

Hypothesized Replication (HR): Numeracy-training participants would demonstrate posttest performance on numeracy tasks consistent with having more improved objective

numeracy as compared to control participants. We note this is a conceptual – not a direct – replication of Park and Brannon's (2013, 2014), as differences exist in the training tasks and the numeracy assessments as we describe above.

Novel Hypothesis 1 (H1): Individuals lower in subjective numeracy would show greater objective numeracy benefits from non-symbolic numeracy training relative to the symbolic numeracy training. Such a result would support non-symbolic training as particularly beneficial to individuals with low confidence in their math ability.

Novel Hypothesis 2 (H2): Numeracy-training participants would demonstrate posttest performance on judgment and decision-making tasks consistent with having greater objective numeracy as compared to control participants. Such results would provide experimental evidence of a causal link between numeracy and decision-making competence.

Novel Hypothesis 3 (H3): The benefit of numeracy training to JDM performance would be mediated by objective numeracy. Such a result would further support a causal link.

## Method

### Participants

Participants in this study were recruited online via Amazon Mechanical Turk (MTurk) over a two-week period, in small cohorts of no more than 50 individuals. We initially recruited 935 individuals via MTurk who began the pretest. We excluded 66 participants who responded from outside of the United States, and an additional 18 who did not identify as native English speakers, leaving an initial sample *n* of 851.

### Incentives

Participants were paid $2.00 to complete the pretest, $3.00 to complete the posttest, and

$3.00 for each training session in which they participated (up to 6 possible), for a maximum total possible reimbursement of $23.00 in the training conditions and $5.00 in the non-intervention control condition. Participants were paid promptly, typically within 24 hours of each session, to encourage retention.

## Procedure

At pretest, participants first completed measures of subjective and objective numeracy. Next, they provided background information, answering questions about their demographics, "growth mindset" (Dweck, 2003) and interest in improving their math skills. They were also asked about their task compliance (i.e., Did they cheat and use a calculator?, Did they pay attention?). After completing this pretest, participants were randomly assigned to one of six conditions: four numeracy training experimental conditions, an active training control, or a non-intervention control (described below). Training participants went on to complete six training sessions on six separate days. Finally, participants completed a posttest consisting of three JDM measures, a repeat of the pretest numeracy measures, two additional numeracy measures evaluating objective numeracy and symbolic number mapping (given at posttest only due to known practice effects, Chesney et al. 2015), and questions about compliance. (See Table 1 for time-line).

### Measures

***Numeracy Measures.*** We measured participants' subjective numeracy at pretest and posttest using the Subjective Numeracy Scale (SNS, Fagerlin et al., 2007), which asks participants eight questions regarding their comfort using numbers (e.g. "How good are you at working with percentages?"). Objective numeracy was measured at pretest and posttest with an Arithmetic task following the procedure described by Park and Brannon (2013, 2014), in which

participants completed as many symbolic addition and/or subtraction problems as possible in 10 minutes (e.g. "(36-214)+202)" ). Additionally, at posttest we gave a version of the Objective Numeracy Scale (ONS, see Weller, Dieckmann, Tusler, Mertz, Burns, & Peters, 2013), based on items provided by Edward Cokely (personal communication, January 21, 2015). This set of seven word problems required participants to calculate probabilities, percentages and proportions, such as "Imagine that we flipped a fair coin 1,000 times. Out of 1,000 flips, how many times do you think the coin would come up as heads?" We assessed participants' symbolic number-mapping ability at posttest via an SMap task from Schley and Peters (2014), in which participants placed 71, 996, 780, 982, 6, 770, 230, 994, 18, 220, and 4 on a 0-1000 number line. Tasks are described in greater detail in the supplement (section 1.1).

***Judgment-and-Decision-Making (JDM) measures*.** Participants completed three JDM measures at posttest-only: a Consistency-in-Risk-Perception task (see Bruine de Bruin, Parker, & Fischhoff, 2007), the Bets task (Peters et al., 2006), and a Framing task (Peters et al., 2006). The Consistency-in-Risk-Perception task is a within-subjects measure where participants estimate the chances of an event happening or not happening over a given time frame (i.e. "What is the probability [nothing will be stolen / that someone will steal something / that someone will break into your home and steal something] from you during the next [year / 5 years]?"). Scores reflect the consistency of judgments with each other. The Bets task and the Frame task are between-subjects measures. The Bets task compares bet-attractiveness judgments made by participants evaluating a bet with a small loss to judgments made by other participants evaluating a similar bet without a loss (i.e. "There are 7 chances out of 36 that you will win the bet and receive $9.00 and 29 chances out of 36 that you will [win nothing / lose 5 cents]"). The Framing task compares ratings of student performance by participants seeing that performance described in a positive

frame to ratings from other participants seeing that same performance described in a negative frame (i.e. “80% correct” vs. “20% incorrect”). Objective numeracy has been related to performance on all three tasks (Gamliel, Kreiner, & Garcia-Retamero, 2016; Peters et al., 2006; Peters & Bjälkebring 2015; Peters, Fennema, & Tiede, 2019; Sinayev & Peters, 2015). Tasks are described in greater detail in the supplement (section 1.2).

**Table 1**.

*Timeline of Procedures*

| Day | Activities |
|---|---|
| Session 1 | 1. Pretest (15-20 min):<br>Numeracy measures:<br>Subjective Numeracy Scale (SNS)<br>Arithmetic task<br>Other questions (e.g., background information and task compliance)<br><br>2. Assignment to condition<br><br>3. Training (if applicable) (20-30 min) |
| Sessions 2-5 | 1. Training (if applicable) (20-30 min, 24-72 hours after the prior session) |
| Session 6 | 1. Training (if applicable) (20-30 min, 24-72 hours after the prior session)<br><br>2. Posttest (20-30 min):<br>JDM tasks:<br>Bets task<br>Consistency in Risk Perceptions<br>Framing<br>Numeracy measures:<br>SNS<br>Arithmetic task<br>Objective Numeracy Scale (ONS)<br>Symbolic Mapping (SMap) task<br>Other questions (e.g., task compliance) |

Note: Training participants were linked to the posttest immediately after their last training session. Non-intervention controls were sent this link when half their recruitment cohort completed training.

**Assignment to condition.** Participants were randomly assigned to one of six possible conditions: one of the five training conditions (detailed below) or a non-intervention control. Participants assigned to the non-intervention control were invited to participate in the posttest in 1-2 weeks. Participants assigned to the training conditions were invited to participate in six

training sessions and the posttest, all to be completed within 1-2 weeks. Participants in all training conditions (including the active memory-training control) were told that they would practice "skills related to math ability" to equalize demand characteristics between intervention conditions. Spatial and math skills have long been linked (Verdine, Irwin, Golinkoff, & Hirsh-Pasek, 2014). Participants who accepted the training invitation were linked immediately to the training website in a new window to complete their first session. They were again given the option to decline participation at this time.

**Training.** Each of the six training sessions typically took 20-30 minutes to complete. Participants were able to start each subsequent training session at any time within a 24- to 72-hour window after they had started the previous session. A reminder email with a link to the training webpage was sent after 24 hours had passed. If participants did not begin the next session within the 24-72 hour window, they were excluded from the study.

In the four numeracy intervention conditions, participants practiced approximate arithmetic following a 2 (addition, subtraction) × 2 (symbolic, non-symbolic) between-subjects design. In particular, they practiced either approximate addition OR approximate subtraction using either symbolic OR non-symbolic numeric stimuli. Otherwise, we followed the training procedures described by Park & Brannon (2013, 2014). In the symbolic condition, stimuli were Arabic numbers. In the non-symbolic condition, the stimuli were presented as arrays of black circles on an off-white background. Circles were randomly distributed in a 200x200 pixel region with actual size dependent on the participants' screen resolution. Circles were not allowed to overlap. The continuous extent of these dot arrays was carefully controlled to prevent area or circle size from being a consistent cue to the numerosity of the set: Total circle area was randomly selected from a pair of possible areas, one twice the size of the other. Individual circle

size was randomized, with minimum possible circle diameter set at 4 pixels while the maximum circle diameter varied with the size of the set ensuring all circles could be drawn in a non-overlapping fashion while maintaining the selected total area."

In the addition conditions, values would fly in from the left and right of the screen and hide behind a grey square. In subtraction conditions, a value would fly in from the left, and a second would fly out from the right. Participants estimated the sum or difference between these values. Participants then compared their estimates to a third value, either by saying if their estimate was greater or less than this value (comparison trials) or by choosing the correct value from a pair where the comparison-value acted as the foil (match trials, see Figure 1). The ratio between the actual sum or difference and this comparison-value became smaller (more difficult) when participants responded correctly, and larger (less difficult) when participants made mistakes. The fifth training condition was an active control following Park and Brannon (2014), in which participants practiced a spatial working-memory task. Participants were asked to repeat, backwards or forwards, a sequence of indicated squares in a 4 × 4 grid. The sequence became longer or shorter as the participants succeeded or failed. The supplement (section 1.3) describes training in greater detail.

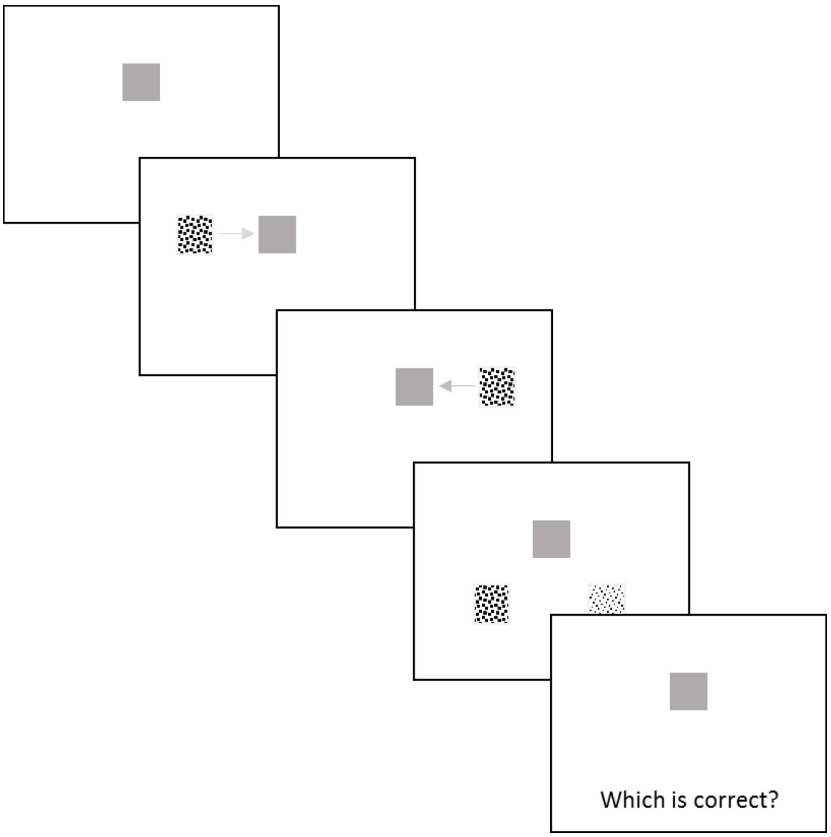


*Figure 1. Cartoon illustrating procedure used in non-symbolic addition "match" training trials. In other "comparison" trials, only the comparison value would be presented, and participants would say if the sum or difference was greater or less than this value.*

## Results

### Tests to Confirm Equivalent Sampling Between Groups:

**Retention was similar among training groups, but different for the non-intervention control.** After excluding individuals outside the USA and non-native English speakers, our eligible recruitment sample size was 851, with 138-144 participants assigned to each group. Of these participants, 48.3% ($N = 411$) were retained over all sessions. The final sample was 51.6% male, 48.4% female; 80.8% white, 7.8% Asian, 7.1% African American or Black, 5.8% Hispanic, 1.0% Native American, 1.0% "other"; mean age 35.3 years, *SD* 10.9, Range 19-69. Lack of retention

was due to declining to participate, failing to complete all six training sessions, failing to complete the posttest, using a calculator in the pretest or posttest arithmetic tasks, or noncompliance on the training tasks (see Table 2). Among the training groups, 86.5% of participants were retained, on average, between each of the six repeated sessions over two weeks (i.e., between sessions 1 and 2, between sessions 2 and 3, etc.). Such retention is considered high (Bartels, 2000; Hansen, 2008; Keith, Tay, & Harms, 2017).

**Table 2**.

*Retention Rates and Attrition Causes Among the Six Between-Subjects Conditions*

| | | # Lost to Attrition by Cause | | | | |
|---|---|---|---|---|---|---|
| Condition | Assigned | Declined | Training Incomplete | Posttest Incomplete | Other Exclusion | Final N |
| Memory | 144 | 29 | 51 | 0 | 2 | 62 |
| Symbolic Addition | 143 | 32 | 42 | 0 | 2 | 67 |
| Non-symbolic Addition | 139 | 35 | 34 | 1 | 5 | 64 |
| Symbolic Subtraction | 138 | 28 | 40 | 0 | 3 | 67 |
| Non-symbolic Subtraction | 144 | 26 | 48 | 0 | 9 | 61 |
| Non-intervention | 143 | 4 | N/A | 43 | 6 | 90 |

Active memory training control retention (43.1%) did not differ from retention across numeracy training interventions (45.9%; $X^2$ (1, $N$ = 708) = .38, $p$ = .537). However, non-intervention participants were more likely to be retained (62.9%) than those assigned to the numeracy training intervention and memory training control conditions (42.4-48.6%) ($X^2$ (1, $N$ = 851) = 14.75, $p$ < .001). These findings suggest likely fundamental selection-effect differences between participants in the non-intervention control and those in the training conditions that result from the unavoidable confound that it is easier to retain participants over fewer sessions. Thus, data from the non-intervention control were excluded from further analysis, leaving 321

active-training participants. Nonetheless, means for the non-intervention participants are provided in tables for comparison purposes.

**Table 3**

*Pretest and Posttest Mean (SE) Scores on Numeracy Measures by Condition*

| Condition | Pretest SNS | Posttest SNS | Pretest Arithmetic | Posttest Arithmetic | Posttest ONS | SMap* |
|---|---|---|---|---|---|---|
| Possible range of scores | 1-6 | 1-6 | 0-96 | 0-96 | 0-7 | 0-193* |
| Memory ($N$ = 62) | 4.63 (0.10) | 4.69 (0.12) | 40.31 (2.04) | 40.94 (2.17) | 3.44 (0.24) | 23.83* (1.59) |
| Symbolic Addition ($N$ = 67) | 4.58 (0.09) | 4.71 (0.09) | 37.94 (1.80) | 39.06 (1.82) | 3.82 (0.20) | 23.12 (1.42) |
| Non-symbolic Addition ($N$ = 64) | 4.68 (0.11) | 4.83 (0.11) | 36.83 (1.94) | 38.06 (2.12) | 3.73 (0.21) | 25.11* (1.51) |
| Symbolic Subtraction ($N$ = 67) | 4.88 (0.09) | 4.93 (0.10) | 42.54 (1.92) | 43.07 (2.06) | 4.04 (0.19) | 23.58 (1.46) |
| Non-symbolic Subtraction ($N$ = 61) | 4.65 (0.11) | 4.74 (0.11) | 39.64 (2.09) | 39.80 (2.06) | 3.87 (0.18) | 25.15* (1.46) |
| Non-Intervention Control ($N$ = 90) | 4.67 (0.07) | 4.77 (0.08) | 37.91 (1.55) | 35.83 (1.51) | 3.57 (0.18) | 28.54* (3.07) |

*Note: SMap scores reflect the mean absolute numeric distance from correct (ADC) of the participants' placement of numbers on the 0-1000 number line, excluding the placement of '71'. We excluded 4 participants whose resulting ADCs were more than 5 SDs above the mean: 1 memory, 1 non-symbolic addition, and 2 non-symbolic subtraction participants. One non-intervention control participant did not complete the SMap task.

## Sensitivity Analysis

We used G*Power version 3.1.9.2 (2014) to determine the sensitivity of our final sample size of 321. The sensitivity analysis showed that with an $N$ of 321, an ANCOVA with 1 covariate (i.e., pretest Arithmetic), 1 degree of freedom in the numerator (i.e., our main contrast between memory training and numeracy training groups), 5 groups, and an α error probability of .05 can detect an effect size $f$ of .157 with 80% power. This corresponds to an $\eta_p^2$ of 0.024. Thus, our study was able to detect effects sizes substantially smaller than the $\eta_p^2$ of 0.132 reported by Park and Brannon (2014).

**Groups were equally numerate at pretest.** Attribution in randomly assigned studies rarely changes results (Chandler & Shapiro, 2016) and random assignment should yield groups equivalent at pretest (Kelly & Maxwell, 2010). Nonetheless, we examined whether the final members of the five active training groups differed in pretest subjective or objective numeracy. A one-way ANOVA revealed no significant condition differences in pretest SNS scores overall ($F(4, 316) = 1.378$, $p = .241$, $\eta_p^2 = .017$) or in pairwise comparisons (Tukey HSD, all $p$s > .190). Similarly, no significant differences emerged in pretest Arithmetic scores overall ($F(4, 316) = 1.303$, $p = .269$, $\eta_p^2 = .016$) or in pairwise comparisons (Tukey HSD, all $p$s > .220). Specific contrasts also revealed no difference between the memory training group and numeracy training groups in pretest SNS ($t(319) = -.621$, $p = .535$) or pretest Arithmetic ($t(319) = .473$, $p = .636$). Bayesian ANOVAs run using JASP 0.14.1 (JASP Team, 2020) following a method described by van den Bergh et al. (2019) confirmed that the Null model was more likely than group differences in pretest SNS ($BF_{10} = .070$) or pretest Arithmetic ($BF_{10} = .061$). Pretest and posttest mean scores are available in Table 3.

## Tests of Hypothesized Replication (HR) of Park and Brannon (2013, 2014)

**Training performance was somewhat similar to Park and Brannon (2013, 2014).** Participants generally followed the same pattern of improvement across the training sessions as seen by Park and Brannon (2013, 2014). They greatly improved between sessions 1 and 2, with little improvement thereafter (see supplement, Table S2). However, we do note some differences. Our memory-training participants only recalled 4.42 item sequences by the end of session 6, whereas Park and Brannon's (2014) participants recalled 5.2 item sequences. Among our non-symbolic arithmetic training groups, participants in the addition condition got down to a mean

discrimination ratio (i.e., difficulty level) of 1.51/1 by the end of session 6 while the non-symbolic subtraction participants only got down to 2.55/1, confirming that the non-symbolic subtraction task was substantially more difficult. In contrast, Park and Brannon's (2014) participants got down to a discrimination ratio of 1.56/1, similar to the present non-symbolic addition participants, while practicing a mix of addition and subtraction trials.

**Partial replication of the benefits of approximate arithmetic training on objective numeracy.** Our participants' arithmetic performance was substantially different than that seen by Park and Brannon (2014). Park and Brannon reported that their non-symbolic arithmetic training participants correctly answered an average of 67.3 exact arithmetic items at pretest, and 81.7 at posttest, with this improvement being significantly larger than that seen in their active controls ($F(3,69) = 3.946$, $p = 0.012$, $\eta_p^2 = 0.132$). In contrast, our participants' group means cluster around 40 at both pretest and posttest (see Table 3). However, although Park and Brannon (2014) measured objective numeracy using the Arithmetic task alone, we measured objective numeracy in two ways: the Arithmetic task (assessed at pretest and posttest) and the ONS task (posttest only). This additional task gave us another measure of objective-numeracy performance between our experimental conditions.

To account for pretest individual differences and these multiple measures, we used SPSS version 25 to conduct a Generalized Estimating Equation (GEE) analysis of posttest objective numeracy with the two subscales - Arithmetic and ONS - treated as repeated measures. This analysis allowed us to simultaneously test whether numeracy training improved objective numeracy versus the active memory-training control and if it had differential effects on the two subscales, while controlling for individual differences in pretest objective numeracy. For ease of comparison, ONS scores and pretest and posttest Arithmetic scores were transformed into

proportion correct; pretest Arithmetic scores were entered as a covariate. We used maximum likelihood estimation with a normal probability distribution, identity link function, and independent correlation matrix to examine the two-way interaction of objective numeracy subscale (posttest Arithmetic vs. ONS) and training (numeracy training vs. memory training).

We only partially replicated Park and Brannon's (2013, 2014) results. Specifically, we found that participants who received numeracy training demonstrated marginally better posttest objective numeracy than memory training control participants (numeracy training: $M = 0.49$, $SE = 0.01$, $N = 259$; memory training: $M = 0.45$, $SE = 0.02$, $N = 62$), Wald $\chi^2(1) = 2.96$, $p = .086$). Participants, on average, had a higher proportion correct on the ONS subscale ($M = 0.52$, $SE = 0.02$) than the Arithmetic subscale ($M = 0.42$, $SE = 0.01$), Wald $\chi^2(1) = 36.19$, $p < .001$, and people who did better on the pretest Arithmetic task also did better on the posttest objective numeracy subscales, *b(se)*= 0.67 (0.04), Wald $\chi^2(1) = 266.70$, $p < .001$). However, the main effects were qualified by a significant interaction of training condition and objective numeracy subscale, (Wald $\chi^2(1) = 4.60$, $p = .032$): The effect of numeracy vs. memory training was significant for the ONS subscale (Wald $\chi^2(1) = 4.57$, $p = .033$), but not for the Arithmetic subscale (Wald $\chi^2(1) < 1$). Similar results were found in an alternative analysis transforming the Arithmetic and ONS to z-scores, (see supplement, section 2.1.1).

We confirmed that posttest Arithmetic scores were not influenced by training condition via a Bayesian ANCOVA that included training condition as a fixed factor and pretest Arithmetic scores as a covariate. It was conducted with JASP 0.14.1 (JASP Team, 2020) following the method described by van den Bergh et al. (2019). Model comparisons indicated that the best model included only pretest Arithmetic as a predictor of posttest Arithmetic scores. The model including both pretest Arithmetic and numeracy training condition ($BF_{10} = 0.153$), the

model including only numeracy training condition ($BF_{10} < .001$), and the null model ($BF_{10} < .001$) were all less likely. (Note: We report all $BF_{10}$s in reference to the best model, whose $BF_{10}$ is fixed at 1). Additional analyses also found no benefits of training on the SNS and SMap measures (see supplement, section 2.1.2).

In order to determine whether this arithmetic benefit occurred both for participants receiving non-symbolic training (the training that most closely replicated that used by Park and Brannon, 2013, 2014) and for our novel symbolic training, we conducted separate ANCOVAs contrasting memory training participants with only *non-symbolic* arithmetic training participants or only *symbolic* arithmetic training participants (again including pretest Arithmetic scores as a covariate). Non-symbolic participants had marginally better posttest ONS-scores than memory participants ($F(1,184) = 3.341$, $p = .069$, $\eta_p^2 = .018$), but showed no differences on the posttest Arithmetic task ($F(1,184) = 0.035$, $p = .853$, $\eta_p^2 = .000$). Symbolic participants showed the same pattern (ONS: $F(1,193) = 4.393$, $p = .037$, $\eta_p^2 = .022$; Arithmetic: $F(1,193) = 0.017$, $p = .896$, $\eta_p^2 = .000$).

We also ran a GEE analysis which found no significant differences among numeracy-training conditions (see supplement, section 2.1.3). We confirmed that there were similar numeracy outcomes between non-symbolic and symbolic numeracy training groups with a Bayesian ANCOVA on numeracy training participants' posttest ONS scores, including non-symbolic vs. symbolic training as a fixed factor, and pretest Arithmetic scores as a covariate (again conducted with JASP 0.14.1; JASP Team, 2020). Model comparisons indicated that the best model included only pretest Arithmetic as a predictor of posttest ONS scores. The model including both pretest Arithmetic and numeracy training condition ($BF_{10} = 0.146$), the model including only numeracy training condition ($BF_{10} < .001$), and the null model ($BF_{10} < .001$) were

all less likely.

Although detectable, the present effect size is substantially smaller than the medium-sized effect ($\eta_p^2 = 0.132$) found by Park and Brannon (2014). It is possible that the smaller effect size with numeracy benefits seen in ONS but *not* in Arithmetic scores (unlike Park and Brannon's [2013, 2014] work) was due to differences in our samples and training. We discuss possible reasons for these differences in the General Discussion. Nevertheless, as some improvements to the training participants' numeracy were detected, we went on to test our novel hypotheses regarding the benefits of numeracy training to JDM performance.

**Tests of Novel Hypotheses**

**Test of H1 (that less subjectively numerate individuals would benefit more from non-symbolic than symbolic training).** We found that Non-symbolic training was particularly likely to yield higher posttest numeracy scores for low subjective-numeracy participants. A GEE analysis found that, as hypothesized in H1, the interaction of symbolic vs. non-symbolic condition and pretest SNS was significant (Wald $\chi^2(1) = 5.42$, $p = .020$). Among participants with lower pretest SNS scores (-1 SD), the effect of symbolic vs. non-symbolic condition was negative albeit non-significant, *b(se)* = −0.03 (0.02), Wald $\chi^2$ (1) = 1.81, $p = .179$. Among participants with higher SNS scores (+1 SD), the effect of symbolic vs. non-symbolic condition was positive, *b(se)* = 0.04 (0.02), Wald $\chi^2(1) = 3.20$, $p = .074$. Higher SNS predicted greater math performance in the symbolic training condition *b(se)* = 0.10 (0.02), Wald $\chi^2(1) = 21.39$, $p < .001$, but this effect was reduced in the non-symbolic training condition *b(se)* = 0.04 (0.02), Wald $\chi^2(1) = 2.93$, $p = .087$. A detailed description of this and additional analyses is available in the supplement (section 2.2).

**Test of H2 (that numeracy training would result in posttest JDM performance consistent with having greater objective numeracy).** Participants responded to three JDM tasks: bets, framing, and risk-perception consistency. However, the anticipated, pre-requisite effects of objective numeracy in the between-subject framing and bets tasks did not replicate. Hence, only the within-subject risk-perception task could be used to test H2. We discuss this decision further in the supplement (sections 2.3, 2.4, & 3).

A GEE analyses showed that, consistent with H2, inconsistencies in risk judgments were less likely among numeracy-training participants (proportion-based analysis: $M = 0.17$, $SE = 0.02$; binary-based analysis: $M = 0.33$, $SE = 0.02$) than memory-training participants (proportion-based analysis: $M = 0.12$, $SE = 0.01$, Wald $\chi^2(1) = 5.33$, $p = .021$; binary-based analysis: $M = 0.42$, $SE = 0.05$, Wald $\chi^2(1) = 3.28$, $p = .070$); consistency in risk perceptions did not differ between the numeracy-training conditions. Analysis details are available in the supplement (section 2.3).

**Test of H3 (that the benefit of numeracy training on judgments would be mediated by objective numeracy).** We were interested in determining whether the effect of numeracy vs. memory training on risk-inconsistency errors was mediated by the numeracy-training effect on ONS scores. (We restricted our analysis to ONS scores because numeracy training did not influence the Arithmetic scores, and thus were not a possible mediator). Thus, we ran a simultaneous regression mediational analysis, details of which are available in the supplement (section 2.5).

Results supported our hypothesis. As expected based on prior analyses, numeracy training predicted higher ONS scores, (both proportion- and binary-based analyses: *b(se)* = 0.07, (0.03), $t = 2.18$, $p = .030$), as did pretest Arithmetic scores, (both proportion- and binary-based

analyses: *b(se)* = 0.50 (0.07), $t = 6.62$, $p < .001$). Greater ONS predicted less risk inconsistency, (proportion-based analysis: *b(se)* = −0.24 (0.03), $t = -7.46$, $p < .001$; binary-based analysis: *b(se)* = −0.51 (0.07), $t = -6.84$, $p < .001$). Neither pretest Arithmetic (proportion based analysis: (*b(se)* = 0.02 (0.05), $p > .500$; binary-based analysis: (*b(se)* = 0.03 (0.11), $p > .750$) nor numeracy training (proportion based analysis: (*b(se)* = 0.04 (0.04), $p = .053$; binary-based analysis: (*b(se)* = −0.06 (0.04), $t = -1.37$, $p = .171$) predicted significantly fewer inconsistencies when ONS was included in the model, although there was a trend in this direction. The Sobel test of the indirect effect indicated successful mediation, with the effect of numeracy training becoming not significant when included in the same model as ONS, (proportion-based analysis: *IE(se)* = −0.02 (0.01), $z = -2.07$, $p = .038$; binary-based analysis: *IE(se)* = −0.03 (0.02), $z = -2.06$, $p = .040$). Additionally, the 95% confidence interval of the indirect effect using bootstrapping with 1,000 resamples (proportion-based analysis: −.04 to −.002; binary based analysis: −.07 to −.003) did not include zero, further indicating significant mediation (Preacher, Rucker, & Hayes, 2007; Shrout & Bolger, 2002; Zhao, Lynch, & Chen, 2010).

### Supplementary Materials and Data Files

Supplementary material, data files, and SPSS analysis scripts are available here: https://osf.io/x75ys/?view_only=15c319943c474de5a98a39236ec822e5

## General Discussion

### Insights into the Benefit of Non-symbolic Training to Objective Numeracy

**Non-symbolic training was as beneficial as symbolic training.** Although we found that non-symbolic training was particularly beneficial for individuals with lower subjective numeracy, overall, non-symbolic and symbolic training were equally beneficial. Specifically,

objective-numeracy outcomes did not differ, on average, between these conditions. This finding surprised us. On the face of it, one might think that training focused on symbolic numbers would be more beneficial to other symbolic tasks than would non-symbolic training. Park and Brannon (2013, 2014) did not test this comparison and, instead, contrasted non-symbolic training only with math-free controls. However, it appears that the choice of non-symbolic or symbolic stimuli was unimportant to the average benefit provided by the approximate-arithmetic-based numeracy training.

**No evidence that training improved symbolic-number mapping or subjective numeracy.** The current results did not support one of Park and Brannon's (2013) suggestions that their non-symbolic training transferred to symbolic arithmetic by improving ANS-acuity which we assessed via symbolic-number mapping in the current study. Our non-symbolic numeracy training did not significantly improve symbolic-number mapping. However, Park and Brannon's alternative explanation remains plausible, namely that symbolic-arithmetic improvement resulted from an unidentified "common cognitive component of mental quantity manipulation" (p. 247, Park & Brannon, 2016) involved in both symbolic and non-symbolic arithmetic. A third possibility is that non-symbolic-arithmetic practice transferred to symbolic arithmetic via associative mappings between non-symbolic and symbolic quantities. In other words, non-symbolic numeric quantities (e.g., :::) are thought to activate their corresponding symbolic values (e.g., "6") and vice versa (Dehaene, 1992). Thus, practicing non-symbolic arithmetic might transfer to symbolic tasks along these pathways (i.e., adding : and :: to get ::: would activate the corresponding symbols in an additive context, thereby reinforcing that 2 + 4 = 6), thus improving math fact knowledge without necessarily improving symbolic-magnitude mapping itself.

The results also do not support the idea that non-symbolic training improved symbolic arithmetic by improving subjective numeracy: No subjective numeracy differences were detected between the numeracy-training and control conditions (see supplement 2.1.2). Of course, if non-symbolic training did not improve symbolic number mapping or subjective numeracy, it could not have benefitted objective numeracy and judgment through these constructs. Taken together, these results indicate that numeracy training benefitted objective numeracy via some other mechanism (e.g., some common cognitive component of mental quantity manipulations or associative mappings between symbolic and non-symbolic quantities). Further study is needed to determine exactly what this mechanism might be.

**Non-symbolic Training May Help Individuals with Less Confidence in Their Math Abilities**

Participants who had lower SNS scores (e.g., who rated themselves as worse with numbers and preferring to use them less) derived directionally more benefit from non-symbolic training, whereas those with higher SNS scores instead performed marginally better when they received training with traditional symbolic numbers. It is important to emphasize that this effect was detected while controlling for pretest objective numeracy, indicating that it was the participants' *beliefs* about their numeric ability—not their objective ability—that was the moderating factor. Additionally, we did not measure math anxiety specifically, but subjective numeracy and math anxiety are related constructs. Thus, this effect may indicate that training with non-symbolic numbers offers a "back door" of sorts to improve numeric ability among less math-confident and possibly less math-anxious individuals. This conjecture makes some sense because anxiety can interfere with the performance of math-anxious individuals, a vicious cycle that reinforces math anxiety (Maloney & Beilock, 2012). Non-symbolic training may allow math skills to be practiced in a less anxiety-producing context, because the absence of symbolic

numbers allows individuals to reap the benefits of math practice without interference from math anxiety. Future replication attempts investigating this issue should include larger samples of math anxious individuals and/or include specific measures of math anxiety.

**A Causal Link Between Objective Numeracy and Risk Judgments**

Numeracy training yielded more consistent risk perceptions, and this benefit was mediated by post-intervention condition differences in objective numeracy (controlling for pre-intervention arithmetic scores). These results indicate that the benefits of numeracy training can extend beyond mathematical paradigms to improved judgments. In addition, training need not be rooted in traditional symbolic calculation. Specifically, approximate-arithmetic training can yield these benefits, using either symbolic or non-symbolic numbers.

The precise mechanism for how numeracy causes these improvements remains unclear (see Reyna et al., 2009). It may be that more objectively numerate individuals: 1) habitually make more numeric comparisons and transformations (Peters et al., 2006; Peters, Fennema, & Tiede, 2019); 2) engage spontaneously in greater deliberation about numeric information (Ghazal, Cokely, & Garcia-Retamero, 2014; Obrecht & Chesney, 2016; Peters et al., 2006); 3) have a more accurate understanding of numeric magnitudes that they use to value numeric information in decisions (Peters, Slovic, Vastfjall, & Mertz, 2008; Schley & Peters, 2014); and/or 4) have adequate efficacy with numbers, enabling them to make consistent judgments based on numeric information (Peters & Bjälkebring 2015; Rolison et al., 2016). We note these explanations are not mutually exclusive. The latter two mechanisms, however, are less likely in the present case given that we found no significant effects of numeracy vs. memory training on symbolic number mapping or subjective numeracy. The specific mechanisms yielding the observed effects might be addressed in future work with tailored questions.

## Limitations and Future Directions

**Differences from Park and Brannon (2013, 2014).** Our results diverge from Park and Brannon's (2013, 2014) in some substantial ways. The benefit to objective numeracy was seen only in the ONS measure and not in the posttest Arithmetic measure, even though Park and Brannon saw benefits in their Arithmetic task. In addition, this benefit was small compared to the medium-sized effect observed by Park and Brannon (2014). On a possibly related note, in training, our non-symbolic subtraction participants did not attain the difficulty levels reached by Park and Brannon's (2014) non-symbolic arithmetic participants, and our memory participants also lagged behind the difficulty levels reached by their counterparts in Park and Brannon's (2014) study. We next consider possible reasons for these differences.

First, our online training may have been less effective than the in-lab training conducted by Park and Brannon (2013, 2014). Also, we did not interleave subtraction and addition training trials: This difference may have reduced the manipulation's efficacy (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Participants in our conditions received training in which they practiced the same skill (e.g., addition or subtraction) repeatedly. Research on educational interventions suggests that "interleaved" practice, in which successive problems need to be solved via different mathematical strategies (e.g., a mixture of arithmetic skills) produces better performance on subsequent evaluations (e.g., Mayfield & Chase, 2002; Rohrer, Dedrick, Hartwig, & Cheung, 2019). By separating our problems into between-group conditions, we may have reduced our ability to find the expected effect on arithmetic problems. This explanation, however, does not address the impact of numeracy vs. memory training on ONS problems. It is also possible that the different fixed problem sets we used at pretest and posttest may not have been equally difficult. However, this difference is not relevant to our critical analyses, which

controlled for pretest scores.

It may be that numeracy vs. memory training encouraged abstract vs. concrete processing, instead of or in addition to the anticipated effect on objective numeracy. Abstract processing is thought to improve performance on word problems (like those in the Objective-Numeracy Scale, but NOT the Arithmetic task) because thinking abstractly encourages people to ignore superfluous details in the narrative component of the problem in order to focus on the key numeric information necessary to solve the problem (Schley & Fujita, 2014). This logic could potentially explain why we saw training condition effects on ONS but not Arithmetic. However, research on fuzzy trace theory (Reyna et al., 2009) suggests that abstract processing should also have increased framing effects, and no effect of training condition on framing effects was seen (see supplement, section 2.3).

Finally, the different results could reflect a difference in our samples. For example, Park and Brannon's participants were primarily college students and may have been more accustomed to taking symbolic math tests than participants in our more diverse internet sample. The fact that Park and Brannon's (2014) non-symbolic arithmetic training participants were able to solve 67.3 arithmetic items on average at pretest, whereas our participants averaged about 40 on a similar test might be demonstrative of such a difference and/or the difference between in-person and online training.

**Alternative explanation for Risk-Judgment-Consistency differences: Possible effects of numeric priming?** We concluded that differences existed in the consistency of Risk judgments between the numeracy and memory training groups that were presumably due to our manipulation of objective numeracy. Our ability to interpret the results was aided by the fact that the intervention improved objective numeracy scores and risk judgments in the absence of any

posttest training-condition differences in subjective numeracy or symbolic number mapping. One might otherwise suspect that participants made more numerically consistent risk judgments simply because numeracy training increased their confidence in their own numeric ability and/or their understanding of values represented by symbolic numbers. Nevertheless, it remains possible that numeracy training primed numeric processing more generally and made people more likely to make use of their numeric skills (Hsee & Rottenstreich, 2004), a change in motivation or cognitive activation rather than in objective ability or confidence *per se*.

**A way to target objective numeracy?** The finding that, relative to memory training, numeracy training improved objective-numeracy scores specifically – and not subjective numeracy or symbolic-number mapping – is itself intriguing. Higher objective numeracy has been related to both subjective numeracy and symbolic number mapping in past studies (e.g., Peters & Bjalkebring, 2015). Thus, one might expect these changes to co-occur with improvements to objective numeracy. The fact that objective numeracy *can* be specifically manipulated is a boon to researchers wishing to explore causal relations with objective numeracy. Manipulations that can target other aspects of numeracy specifically would be similarly useful. Such interventions would allow researchers to investigate possible causal relations between JDM performance and other aspects of numeracy (e.g., subjective numeracy, symbolic-number mapping, ANS acuity).

## Conclusions

Although our effects were smaller than those found by Park and Brannon (2013, 2014) and the specific effects found were different, we confirmed that practice with approximate arithmetic yielded benefits to objective numeracy. Moreover, we found that this training benefitted objective numeracy specifically; it did not improve other components of numeracy

(i.e., symbolic number mapping) relative to controls. Thus, this training may serve as a useful tool for researchers wishing to investigate causal relations with objective numeracy in the future. Additionally, it appears that non-symbolic training is, on average, as beneficial as symbolic training, and it is directionally more beneficial for individuals with lower subjective numeracy (i.e., those who report having poorer numeric skills). These results suggest that non-symbolic arithmetic is a potential avenue of intervention for people with low math self-efficacy or math anxiety. Finally, we uncovered initial experimental evidence that the observed link between objective numeracy and risk-consistency judgments was, in fact, causal. More numerate individuals make better use of numbers when judging risks *because* they are more objectively numerate: Increasing individuals' objective numeracy yielded more normative judgments.

**References**

Adelman, C. (2006). The toolbox revisited: Paths to degree completion from high school through college. *U.S. Department of Education.*

Ashcraft, M.H. (2002). Math anxiety: Personal, educational, and cognitive consequences, *Current Directions in Psychological Science*, *11*, 181–185. doi: 10.1111/1467-8721.00196

Bartels, L. M. (2000) Panel effects in the American national election studies. *Political Analysis 8*(1), 1-20.

Barton, A., Cokely, E.T., Galesic, M., Koehler, A., & Haas, M. (2009). Comparing risk reductions: On the dynamic interplay of cognitive strategies, numeracy, complexity, and format. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (pp.2347-2352). Austin, TX: Cognitive Science Society.

Bateman, I., Dent, S., Peters, E., Slovic, P., & Starmer, C. (2007). The affect heuristic and the attractiveness of simple gambles. *Journal of Behavioral Decision Making, 20,* 365-380.

Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *Journal of Counseling Psychology*, *25*(5), 441-448. doi: 10.1037/0022-0167.25.5.441

Betz, N. E., & Hackett, G. (1981). The relationship of career-related self-efficacy expectations to perceived career options in college women and men. *Journal of Counseling Psychology, 28(5)*, 399-410. doi: 10.1037/0022-0167.28.5.399

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology, 92(*5), 938-

956. doi: 10.1037/0022-3514.92.5.938

Bynner, A. J., & Parson, S. (2009). Insights into basic skills from a UK longitudinal study. In S. Reder & J. Bynner (Eds.), *Tracking adult literacy and numeracy skills: Findings from longitudinal research* (pp. 27–58). London, England: Routledge.

Chandler, J. & Shapiro, D. (2016) Conducting Clinical Research Using Crowdsourced Convenience Samples. *Annual Review of Clinical Psychology 12,* 53-81.

Chen, Q. & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica, 148,* 163-172. doi: 10.1016/ j.actpsy.2014.01.016.

Chesney, D. L., Bjälkebring, P. & Peters, E. (2015). How to estimate how well people estimate: Evaluating measures of individual differences in the approximate number system. *Attention, Perception, & Psychophysics, 77,* 2781–2802. doi: 10.3758/s13414-015-0974-6

Chesney, D. L. & McNeil, N. M. (2014). Activation of operational thinking during arithmetic practice hinders learning and transfer. *The Journal of Problem Solving, 7*, Article 4. doi: 10.7771/1932-6246.1165

Chesney, D.L., & Peters, E. M. (May, 2015). *Math Priming Improves Judgments.* Poster presented at the annual convention for the Association for Psychological Sciences, New York, NY.

Chesney, D. L. & Obrecht, N. A. (2012). Statistical judgments are influenced by the implied likelihood that samples represent the same population. Memory and Cognition, 40, 420-433. doi:10.3758/s13421-011-0155-3

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1-42. doi: 10.1016/0010-

0277(92)90049-N

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. Psychological Science in the Public Interest, 14(1), 4–58. doi: 10.1177/1529100612453266

Dweck, C. S. (2003) Ability conceptions, motivation and development, in: L. Smith, C. Rogers & P. Tomlinson (Eds) *Development and motivation: joint perspectives, British Journal of Educational Psychology Monograph Series II, 2,* 13–27.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the Subjective Numeracy Scale. *Medical Decision Making, 27*(5), 672-680. doi: 10.1177/0272989X07304449

Gamliel. E., Kreiner, H., & Garcia-Retamero, R. (2016). The moderating role of objective and subjective numeracy in attribute framing. *International Journal of Psychology, 51(2),*109-16. doi: 10.1002/ijop.12138.

Ghazal, S., Cokely, E. & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making, 9(1),* 15–34.

Hackett, G., & Betz, N. E. (1981). A self-efficacy approach to the career development of women. *Journal of Vocational Behavior, 18,* 326-336. doi: 10.1016/0001-8791(81)90019-1

Hansen, J. (2008). Panel Surveys. In W. Donsbach & M. W. Traugott (Eds.), *The Sage Handbook of Public Opinion Research* (330-339). Los Angeles, CA, Sage Publications.

Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General, 133*(1), 23–30. doi:

10.1037/0096-3445.133.1.23

JASP Team (2020). JASP (Version 0.14.1) [Computer software]

Kaufman, E. L., Lord, M.W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62*, 498–525. doi: 10.2307/1418556

Keith, M. G., Tay, L., & Harms, P. D. (2017) Systems Perspective of Amazon Mechanical Turk for Organizational Research: Review and Recommendations. *Frontiers in Psychology, 8,* 1359.

Kelley, K., & Maxwell, S. E. (2010). Multiple regression. In G. R. Hancock & R. O. Mueller (Eds.), *Quantitative methods in the social and behavioral sciences: A guide for researchers and reviewers.* Erlbaum.

Klaczynski, P. A. (2014). Heuristics and biases: interactions among numeracy, ability, and reflectiveness predict normative responding. *Frontiers in Psychology*, *5*, 665-677.

Läg, T., Bauger, L., Lindberg, M., & Friborg, O. (2014). The role of numeracy and intelligence in health-risk estimation and medical data interpretation. *Journal of Behavioral Decision Making, 27*(2*),* 95-108. doi: 10.1002/bdm.1788

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences, 16,* 404-406. doi:10.1016/j.tics.2012.06.008

Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of applied behavior analysis*, *35*(2), 105–123. doi:10.1901/jaba.2002.35-105

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

National Research Council. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Obrecht, N. A., & Chesney, D. L. (2013). Sample representativeness affects whether judgments are influenced by base rate or sample size. *Acta Psychologica, 142,* 370-382. doi:10.1016/j.actpsy.2013.01.012

Obrecht, N. A., & Chesney, D. (2016). Prompting deliberation increases base-rate use. *Judgment and Decision Making, 11*(1), 1. http://journal.sjdm.org/15/15811/jdm15811.pdf

Obrecht, N. A., & Chesney, D.L. (2018). *Tasks that prime deliberative processes boost base rate use.* In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), Proceedings of the 40th Annual Conference of the Cognitive Science Society (2180-2185). Austin, TX: Cognitive Science Society.

Park, J., & Brannon, E.M. (2013). Training the approximate number system improves math proficiency. *Psychological Science, 24*, 2013-2019. doi: 10.1177/0956797613482944

Park, J., & Brannon, E.M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition, 133*, 188-200. doi: 10.1016/j.cognition.2014.06.011

Park, J. & Brannon, E.M. (2016). How to interpret cognitive training studies: A reply to Lindskog & Winman. *Cognition, 150*, 247-251. doi: 10.1016/j.cognition.2016.02.012

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development, 6*(4)*,* 449-468. doi: 10.1016/0885-2014(91)90049-J

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions.

*Current Directions in Psychological Science*, 21, 31-35. doi: 10.1177/0963721411429960

Peters, E. & Bjälkebring, P. (2015). Multiple numeric competencies: When a number is not just a number. *Journal of Personality and Social Psychology, 108*(5), 802-822. doi:10.1037/pspp0000019

Peters, E., Fennema, M., & Tiede, K.E. (2019). The loss-bet paradox: Actuaries, accountants, and other numerate people rate numerically inferior gambles as superior. *Behavioral Decision Making, 32*(1), 15-29. doi.org/10.1002/bdm.2085

Peters, E. & Levin, I. P. (2008). Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. *Judgment and Decision Making*, *3*, 435–448. http://journal.sjdm.org/8407/jdm8407.pdf

Peters, E., Meilleur, L., & Tompkins, M. K. (2013). *Numeracy and the Affordable Care Act: Opportunities and challenges*. Retrieved from http://www.iom.edu/~/media/Files/Activity%20Files/PublicHealth/HealthLiteracy/Commissioned-Papers/Numeracy-and-the-Affordable-Care-Act-Opportunities-and-Challenges.pdf

Peters, E., Shoots-Reinhard, B., Tompkins, M. K., Schley, D., Meilleur, L., Sinayev, A., Tusler, M., Wagner. L., Crocker, J. (2017). Improving numeracy through values affirmation enhances decision and STEM outcomes. *PLoS ONE, 12(7).* Article ID e0180674.

Peters, E. Slovic, P., Västfjäll, D. & Mertz, C. K. (2008) Intuitive numbers guide decisions. Judgment and Decision Making, 3, 619-635. Available at http://journal.sjdm.org/8827/jdm8827.pdf

Peters, E., Tompkins, M.K., Knoll, M., Ardoin, S.P., Shoots-Reinhard, B., & Meara, A.S. (2019).

Despite high objective numeracy, lower numeric confidence relates to worse financial and medical outcomes. *Proceedings of the National Academy of Sciences (PNAS),* doi.org/10.1073/pnas.1903126116.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science, 17*(5), 407-413. doi: 10.1111/j.1467-9280.2006.01720.x

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*, 422-445. doi: 10.1080/00273170701341316

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. doi: 10.1037/a0017327

Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C.-N. (2019). A randomized controlled trial of interleaved mathematics practice. *Journal of Educational Psychology*. doi:10.1037/edu0000367

Rolison, J. J., Morsanyi, K., & O'Connor, P. A. (2016). Can I count on getting better? Association between math anxiety and poorer understanding of medical risk reductions. *Medical Decision Making, 36,* 876-886. doi: 10.1177/0272989X15602000

Schley, D. R., & Fujita, K. (2014). Seeing the math in the story: On how abstraction promotes performance on mathematical word problems. *Social Psychological and Personality Science, 5*(8), 953-961. doi: 10.1177/1948550614539519

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

10.1177/0956797613515485

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies. *Psychological Methods, 7*, 422-445. doi: 10.1037//1082-989X.7.4.422

Siegler, R., & Opfer, J. (2003). The Development of Numerical Estimation Evidence for Multiple Representations of Numerical Quantity. *Psychological science, 14,* 237-43. doi: 10.1111/1467-9280.02438.

Sinayev, A., & Peters, E. (2015). Cognitive reflection versus calculation in decision making. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2015.00532

Smith, J. P., McArdle, J. J., & Willis, R. (2010). Financial decision making and cognition in a family context. *The Economic Journal, 120,* F363– F380. doi:10.1111/j.1468-0297.2010.02394.x

Taves, E. H. (1941). Two mechanisms for the perception of visual numerousness. *Archives of Psychology, 37(Whole No. 265)*, 1–47.

van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E., Derks, K., … Wagenmakers, E. (2019). A Tutorial on Conducting and Interpreting a Bayesian ANOVA in JASP. https://doi.org/10.31234/osf.io/spreb

Verdine, B. N., Irwin, C. M., Golinkoff, R.M., Hirsh-Pasek, K. (2014). Contributions of executive function and spatial skills to preschool mathematics achievement. *Journal of Experimental Child Psychology, 126,* 37–51. doi: 10.1016/j.jecp.2014.02.012

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. doi:10.1002/bdm.1751

Zhao, X., Lynch, J., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths

about Mediation Analysis. *The Journal of Consumer Research, 37*, 197-206. doi: 10.1086/651257