Revisiting the links between numeracy and decision making: Replication of Peters et al. (2006) with an extension examining confidence

Minrui Zhu  
ORCID: 0000-0002-1316-7649  
University of Hong Kong  
[kirkzhu@connect.hku.hk](mailto:kirkzhu@connect.hku.hk) / [kirkzhu9905@gmail.com](mailto:kirkzhu9905@gmail.com)

^Gilad Feldman  
OCRID: 0000-0003-2812-6599  
University of Hong Kong  
[gfeldman@hku.hk](mailto:gfeldman@hku.hk) / [giladfel@gmail.com](mailto:giladfel@gmail.com)

^Corresponding author

## Author bios:

Minrui Zhu was a thesis student at the University of Hong Kong during the academic year 2021-2.

Gilad Feldman is an assistant professor with the University of Hong Kong psychology department. His research focuses on judgment and decision-making.

## Declaration of Conflict of Interest:

The author(s) declared no potential conflicts of interests with respect to the authorship and/orpublication of this article.

## Financial disclosure/funding:

The author(s) received no financial support for the research and/or authorship of this article.

## Authorship declaration:

Minrui Zhu conducted the replication as part of his thesis in psychology.

Gilad Feldman guided and supervised each step in the project, (later: conducted the pre-registrations, ran data collection), and edited the manuscript for submission.

## Corresponding author

Gilad Feldman, Department of Psychology, University of Hong Kong, Hong Kong SAR; [gfeldman@hku.hk](mailto:gfeldman@hku.hk) ; 0000-0003-2812-6599

## Rights:

CC BY or equivalent license is applied to the AAM arising from this submission. ([clarification](https://bit.ly/rrs-primer))

## Contributor Roles Taxonomy

In the table below, employ CRediT (Contributor Roles Taxonomy) to identify the contribution and roles played by the contributors in the current replication effort. Please refer to <https://www.casrai.org/credit.html> for details and definitions of each of the roles listed below.

|  |  |  |
| --- | --- | --- |
| **Role** | **Minrui Zhu** | **Gilad Feldman** |
| Conceptualization | X | X |
| Pre-registration | X |  |
| Data curation |  | X |
| Formal analysis | X |  |
| Funding acquisition |  | X |
| Investigation | X |  |
| Pre-registration peer review / verification |  | X |
| Data analysis peer review / verification |  | X |
| Methodology | X |  |
| Project administration |  | X |
| Resources |  | X |
| Software | X |  |
| Supervision |  | X |
| Validation |  | X |
| Visualization | X |  |
| Writing-original draft | X |  |
| Writing-review and editing |  | X |

# Abstract

**[IMPORTANT:   
Results were written in past tense using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection.]**

Numeracy is individuals’ capacity to understand and process basic probability and numerical information required to make decisions. We conducted a pre-registered replication and extension of Peters et al. (2006) examining associations between numeracy and positive-negative framing (Experiment 1), frequency-percentage framing (Experiment 2), ratio effect (Experiment 3), and loss vs. no-loss (Experiment 4). We collected data with an online US American Amazon Mechanical Turk sample (*N* =850). Our replication [failed to find/found] support for the original findings regarding associations between numeracy and four decision-making effects: [summary effect sizes+CIs will be added here]. Extending the replication, we [found/failed to find] support for an association between numeracy and confidence [summary effect sizes+CIs will be added here]. Materials, data, and code are available on: <https://osf.io/4hjck/>.

*Keywords:* Numeracy, judgment and decision making, registered report, replication, framing effect, confidence

# PCIRR-Study Design Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Question | Hypothesis | Analysis plan | Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis | Interpretation given different outcomes | Theory that could be shown wrong by the outcomes |
| What is the relationship between numeracy and positive-negative framing effects | Higher numeracy is associated with weaker positive-negative framing effect | Mixed ANOVA  Correlations | Our strategy for all replicated studies:  1. We keep the statistical method of the original paper as it treats numeracy as dichotomized.  2. We treat numeracy as a continuous variable, therefore adapt correlation. | Based on the criteria used by Lebel et al. (2019)  We examine the replicability of the findings of Peters et al. (2006), and support for our suggested extensions. | Attribute framing effect |
| What is the relationship between numeracy and percentage and frequency effects | Higher numeracy is associated with weaker frequency-percentage effects | Factorial ANOVA  Correlations | Frequency-percentage framing effect |
| What is the relationship between numeracy and ratio bias  What is the relationship between numeracy and affect precision | Higher numeracy is associated with more optimal choices in competing affective decisions.  Higher numeracy is associated with higher affective precision, in competing affective decisions. | Chi-square test  Independent t-test  Correlations | Deliberate-experiential thinking modes  Ratio bias |
| What is the relationship between numeracy and affective precision and affect in probabilities and numerical comparisons | Higher numeracy is associated with higher affective precision in probabilities and numerical comparisons.  Higher numeracy is associated with greater affect in probabilities and numerical comparisons. | Factorial ANOVA  Independent t-test  Correlations | The highly numerate will focus more on details of numbers and draw more affective meanings.  Bets effect |
| What is the relationship between objective numeracy and confidence under specific conditions? | The highly numerate is related to higher subjective confidence | Correlations | Associations with subjective confidence and objective numeracy |

*Note*. For the sampling plan please see power analysis in the methods section.

# Revisiting the links between numeracy and decision making: Replication of Peters et al. (2006) with an extension examining confidence

## Numeracy

Decisions involving numbers, math, and statistics are common, and people rely heavily on their ability to accurately interpret, think about, and act on them. Numeracy is defined as the individuals’ capacity to understand and process basic probability and numerical information required to make decisions. Research by Peters et al. (2012) demonstrated that numeracy is a predictor of behavior in judgment and decision-making tasks.

We embarked on a direct replication of Peters et al. (2006) with two primary goals. Our first goal was to conduct an independent replication of the associations between numeracy and four decision-making paradigms. Our second goal was to examine an extension regarding the role of confidence (or, subjective numeracy).

We begin by introducing the literature on numeracy and various decision making biases examined in the chosen article for replication - Peters et al. (2006). We provide a brief overview of the decision-making paradigms, in relation to numeracy. We then discuss our chosen target article, summarize its hypotheses and findings, and introduce our extension on the relationship between confidence, numeracy, and decision-making.

## Attribute framing and numeracy

Framing effect is a well-established phenomenon in psychology and behavioral economics, in which decisions are influenced by the way information is presented, such as variations in valence - positive versus negative framing (Tversky & Kahneman, 1981).

Attribute framing is a type of framing effect and relates to the labeling of a particular attribute of an object or an event. For instance, ground beef with 75%-25% meat-fat ratio could be presented as “75% lean” or “25% fat”. Levin and Gaelth (1988) found that people evaluated beef under the “% lean” framing more positively than in the “% fat” framing. Such framing effects have received empirical support by many follow-up studies (e.g., Freling et al., 2014; Piñon & Gambara, 2005).

Attribute framing is related to how people understand and process numerical concepts, suggesting a possible link between numeracy and framing effects. Some studies found that less numerate people were more susceptible to framing effects (including attribute framing) (Choi et al., 2011; Gamliel et al., 2016; Gamliel & Krenier, 2017). For instance, Gamliel and Kreiner (2017) demonstrated the relationship between numeracy and attribute framing bias: students with lower numeracy rated a university course higher if presented with success rates compared to failure rates. They suggested that decision makers with lower numeracy rely more heavily on “non-numerical information”, whereas those with high numeracy pay more attention to numerical information and attain greater accuracy with numbers. Therefore, lower numeracy may be associated with stronger polarization due to the positive or negative valence of framing presentations.

Peters (2012) suggested that highly numerate individuals have the capacity to go beyond the specific numerical information and understand underlying relational information. For example, when a positive outcome is presented as 75% success rate, the highly numerate are more likely to also infer the complementary proportion of the failure rate of 25%, with similar logic for when the failure rates are presented and success rates are inferred. Therefore, the argument in relation to numeracy was that framing bias is attenuated when one is capable of grasping and processing both the positive and the negative information in a decision.

## Frequency-percentage effect and numeracy

Frequency-percentage effect (or “frequency effect”) is the phenomenon that decision making changes when the numbers are presented in forms of frequency (e.g., 10 out of 100) compared to percentage (e.g., 10%) (Gigerenzer, 1991; Hill & Brase, 2012).

Those higher on numeracy seem less likely affected by whether the number is represented in frequency or in percentage (Dickert et al., 2011; Hill & Brase, 2012; Peters et al., 2011). For instance, Peters (2011) tested the relationship among patients. They informed patients of the side-effects of a medication in either frequency or percentage formats (i.e., 10 out 100 versus 10%) and then asked them to rate risk levels. They found that the less numerate were more likely to perceive the medication as less risky when presented in percentage format than in a frequency format. The possible mechanisms could be similar to those we previously discussed regarding the framing effect. Those higher on numeracy may be able to better understand the frequency and probability information as the same mathematical quantity (Hill & Brase, 2012; Peters. 2012)

## Ratio bias and numeracy

Ratio bias (or numerosity effect) is the phenomenon that people tend to focus on absolute numbers rather than on probabilities (Peters et al., 2008; Reyna et al., 2009). For example, people are more likely to choose to select from a sample with a relatively large numerator/large denominator (e.g. 9 in 100) rather than the preferred odds yet relatively smaller numerator/small denominator (1 in 10).

Reyna and Brainerd (2008) separated ratio bias into a heuristic ratio bias (i.e., identical probabilities in the two samples) and a non-optimal ratio bias (i.e., higher probabilities but smaller absolute numerator or lower probabilities but greater absolute numerator). One classic heuristic ratio bias example was from the study of Miller et al. (1989). Children randomly choose a cookie from one of two cookie jars, one containing 1 chocolate chip and 19 oatmeal cookies and the other containing 10 chocolate chips and 190 oatmeal cookies. The probabilities of having a chocolate chip are the same, yet Miller et al. (1989) found that children preferred to choose from the later one, with the larger numbers.

Peters et al. (2006) demonstrated that lower numeracy was associated with less optimal choices in ratio related decisions.

## Affect and numeracy

Two modes of processing information appear to be affective-experiential and deliberative and are also known as the dual process model (Kahneman, 2003; Sloman, 1996). The model suggests that affective-experiential mode produces thoughts and feelings in a relatively effortless and spontaneous manner, whereas deliberative mode requires conscious reason-based and analytical thinking. Affect may provide information about the goodness and badness of an option and might as a consequence influence further choice processes.

Numeracy has been argued as moderating the association between affect and decision-making (Traczyk & Fulawake, 2016; Rottension & Hsee, 2001), with the potential of aiding decision making, yet may sometimes lead to number overuse and worsen decisions (Peters & Bjalkerbring, 2015; Pachur & Galesic, 2013). Those with higher numeracy seem to draw more precise affective information, then form relevant risk perception, and use that information in making related decisions.

In a demonstration of the possible advantages, Petrova et al. (2014) conducted a study about decision making regarding camera insurance. They found that participants with higher numeracy reported greater negative emotions to 90% chance of losing camera compared to 50% chance. In addition, they were willing to pay more on insurance against the loss when the loss probability was higher. By contrast, participants with lower numeracy seemed less sensitive to the two probabilities levels.

However, there are possible nuances and unintended side-effects to drawing precise numerical information, depending on the defined desired outcome. For example, Kleber et al. (2013) conducted research on donations and they found that numeracy was associated with donation behavior, with the more numerate focusing on projects with the greatest proportion of recipients, whereas those lower on numeracy tended to donate more with increases in both the number of recipients and the total number of people in need.

## Choice of study for replication

We chose the article by Peters et al. (2006) as the target for replication based on the following factors: its impact and potential for improvement on methodological limitations in the original studies.

The article has had much impact on scholarly research in the area of social psychology and judgment and decision making. At the time of writing (March, 2022), there were 1360 Google Scholar citations of the article. In addition, Peter et al. (2006)'s work had important practical implications especially in the domains of medical decision making (Reyna et al., 2009; Okamoto et al., 2012) and financial decision making (Estrada-Mejía et al., 2016; Traczyk, et al., 2018).

We reached out to the authors to request assistance with the original materials, and to try and assess any published and ongoing replication work. They indicated most of the original materials have been lost to time, yet kindly referred us to some of the extensive follow-up literature with conceptual replications and related materials, from which we were able to reconstruct most of the studies. We have also learned of other attempts at a replication of the broader numeracy literature in other languages (Polish) and have been in touch with their authors to coordinate efforts. To our knowledge, there are no published direct close replications of the target article’s studies.

Examining the studies, we believe a direct replication is especially relevant given the low power and some of the statistical method choices. Their Studies 1-4 had 100, 46, 46, and 171 participants, respectively, which may seem low, especially given the interaction and supplementary analyses. Furthermore, the methods employed dichotomizing of the continuous numeracy scale, which we thought could be improved by analyzing as the intended continuous measure, and may allow for more accurate insights and conclusions.

We therefore aimed to revisit the classic phenomenon to examine the reproducibility and replicability of the findings with independent replications. We followed recent growing recognition of the importance of reproducibility and replicability in psychological science (e.g., Brandt et al., 2014; Open Science Collaboration, 2015; van‘t Veer & Giner-Sorolla, 2016; Zwaan et al., 2018) and embarked on a well-powered pre-registered very close replication of Peters et al. (2006).

## Hypotheses and findings in target article

Peters et al. (2006) conducted four studies and we aimed to replicate all of them with needed adjustments and collected in a single data collection, with the experiments displayed in a random order (more on that in the methods section). Below we review the findings in each of the target’s studies. We summarized the target’s hypotheses and our extension’s hypothesis in Table 1.

Table 1

*Summary of hypotheses of replication and extension*

|  |  |  |
| --- | --- | --- |
| Study | Hypothesis | Description of hypothesis |
| 1 | 1(original) | The less numerate show a stronger framing effect than the highly numerate. |
|  | 1 (extension) | Higher numeracy is associated with weaker positive-negative framing effects. |
| 2 | 1 (original) | The less numerate are affected more by the frequency-percentage effect than the highly numerate. |
|  | 1 (extension) | Higher numeracy is associated with weaker frequency-percentage effects. |
| 3 | 1 (original) | The less numerate make less optimal choices in competing affective decisions than the highly numerate. |
|  | 1 (extension) | Higher numeracy is associated with more optimal choices in competing affective decisions.. |
|  | 2 (original) | The less numerate have lower affective precision in competing affective decision than the highly numerate. |
|  | 2 (extension) | Higher numeracy is associated with higher affective precision in competing affective decisions. |
| 4 | 1 (original) | The less numerate make more optimal choices in probabilities and numerical comparisons than the highly numerate. |
|  | 1 (extension) | Higher numeracy is associated with less optimal choices in probabilities and numerical comparisons. |
|  | 2 (original) | The less numerate have lower affective precision in probabilities and numerical comparisons than the highly numerate. |
|  | 2 (extension) | Higher numeracy is associated with higher affective precision in probabilities and numerical comparisons. |
|  | 3 (original) | The less numerate have less affect in probabilities and numerical comparisons than the highly numerate. |
|  | 3 (extension) | Higher numeracy is associated with greater affect in probabilities and numerical comparisons. |
| *Extension: Confidence* | | |
| 1, 2, 3, 4 | 1 | Numeracy is positively associated with confidence. |

*Note*. For each of the hypotheses we reframed the hypotheses deduced from the conclusions in the original article from a dichotomy (high numerate versus low numerate, labeled as “original”) to a continuous association (higher numeracy is associated with…, labeled as “extension”).

### Study 1: Numeracy and positive-negative framing

Study 1 sought to examine the relationship between numeracy and attribute framing. They hypothesized that participants with low numeracy are more likely to be affected by attribute framing.

To test this, they recruited participants through campus newspapers. Participants first answered the numeracy scale developed by Lipkus et al. (2001). Then, they rated the quality of five psychology students' work. Participants were randomly assigned to positive or negative framing conditions. For instance, Emily received either 74% correct or 26% incorrect on her exam.

Peters et al. (2006) dichotomized numeracy to high numerate (9-11 items correct) and low numerate (2-8 items correct). To test the hypothesis, they used a mixed ANOVA. They reported that higher numerate participants were less susceptible for framing bias (*f* = 0.25, 90% CI [0.00, 0.15]).

### Study 2: Numeracy and frequency-percentage framing

Study 2 aimed to examine the relationship between numeracy and percentage-frequency framing effect. They hypothesized that participants with low numeracy are more likely to be affected by frequency-percentage effect. To test this, they recruited university students from a psychology course. Participants read the mental-patient scenario in either a frequency or percentage format and rate that the risk level of that patient who would harm someone. They ran a factorial ANOVA and found that low numerate rated lower in the percentage condition than frequency condition whereas the high numerate rated both conditions similarly (*f* = 0.31, 90% CI [0.00, 0.25]).

### Study 3: Numeracy, affect, and ratio bias

Study 3 intended to explore the association between numeracy and ratio bias as well as numeracy and the influence of affective information. They hypothesized that numeracy is associated with more optimal choices, evoking less affect and higher affective precision.

To test this, they recruited university students from a psychology course. Participants from Studies 2 and 3 were the same group. Participants read the scenario about two bowls, Bowl A with affectively appealing description but less objectively favorable outcome and Bowl B with less appealing description but better results. Participants rated their preference for a bowl and selected one. After indicating the preference and choice, they rated affect regarding Bowl A.

They used a chi-square test to examine participants’ choices of two bowls and found that the less numerate were more likely to choose Bowl A (*φ* = 0.767) and that the highly numerate also had higher preference for Bowl B (*d* = -0.74, 95% CI [-1.33, -0.13]). In addition, the high numerate reported higher affect precision on Bowl A (*d* = 0.77, 95% CI [0.17, 1.36]). The study reported no support for differences in feelings (*d* = 0.46).

### Study 4: Numeracy and bet type (loss vs. no-loss)

Study 4 examined the relationship between numeracy and affect in probabilities and numerical comparisons. They hypothesized that numeracy is associated with affect arousal and affective precision.

To test this, they recruited volunteers from a subject’s pool of psychology department. Participants read the scenario about a bet with 7/36 chance to win and 29/36 chance to win nothing or a bet with 7/36 chance to win but 29/36 chance to lose 5 cents. The possible bets were visualized with a roulette wheel. Participants evaluated the attractiveness of the bet and their affect precision and affect, using the same scales as in Study 3.

They employed a factorial ANOVA and an independent samples t-test. They found that those high on numeracy rated the loss bet as more attractive in loss bet condition, whereas participants with low numeracy rated two conditions the same on average (*f* = 0.23 90% [0.01, 0.11]). With respect to affect precision, participants with high numeracy had clearer feelings about the bets than those with low numeracy. The high numerate also reported more positive affect in the loss condition than in the no-loss condition, whereas there were weaker differences for the low numerate (*f* = 0.20, 90% [0.01,0.10]). Peters (2020) summarized such findings as “bets effect” in her book and we will use this term as well in this article.

We summarized the findings in the target article in Table 2.

Table 2

*Summary of original findings in the target article*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S** | **Factors** | **E** | **Effect** | **% CIs** | **CIL** | **CIH** |
| 1 | Numeracy and framing effect | *f* | 0.25 | 90% | 0.00 | 0.15 |
| 2 | Numeracy and frequency-percentage effect | *f* | 0.31 | 90% | 0.00 | 0.25 |
| 3 | Numeracy and bowl choice | *φ* | 0.77 | 95% | / | / |
|  | Numeracy and preference for bowls | *d* | -0.74 | 95% | -1.33 | -0.13 |
|  | Numeracy and affect precision | *d* | 0.77 | 95% | 0.17 | 1.36 |
|  | Numeracy and affect | *d* | 0.46 | 95% | / | / |
| 4 | Numeracy and attractiveness of bet | *f* | 0.23 | 90% | 0.01 | 0.11 |
|  | Numeracy and affect precision | *f* | / | / | / | / |
|  | Numeracy and affect | *f* | 0.20 | 90% | 0.01 | 0.10 |

*Note*. CIL = lower bounds for CIs. CIH = higher bounds of CIs. We report 90% CIs for ANOVA etq-squared given the effect size is always positive.

## Extension: Confidence

We aimed to extend the replication by examining decision-making confidence. Confidence regarding a decision involving statistics may be considered as a measure of subjective numeracy or numeric self-efficacy, concerning how confident people are in their ability to understand numeric information and use mathematical concepts (Peters, 2020, p.5). We discuss two rationales for this extension.

First, there are mixed findings regarding the association between subjective and objective numeracy. A body of research illustrates that subjective numeracy is positively associated with objective numeracy (Garcia et al., 2015; Nelson et al., 2013; Peters et al., 2019). According to the Health Information National Trends Survey conducted by Nelson et al. (2013), participants who regarded themselves as high in subjective numeracy had higher correction rates of objective numeracy questions. Another recent study done by Rolison et al. (2020) illustrated that individuals with higher objective numeracy were more likely to have correct answers in health risk comprehension questions. However, some research found no support for such an association and people with low objective numeracy sometimes deem themselves as highly numerate (Liberali et al., 2012; Gamiel et al., 2016; Peters et al., 2019). For instance, Peters et al. (2019) reported that the objective numeracy sometimes mismatches subjective confidence: 31% participants with high numeracy but low confidence and 44% participants with low numeracy but high confidence.

Second, most current studies measure trait subjective numeracy with self-report questionnaires and two frequently-used scales are Subjective Numeracy Scale developed by Fagerlin et al (2007) and STAT-Confidence Scale developed by Woloshin et al. (2005). Self-report questionnaires target participants' traits or general impressions about their numeracy competence and preference for numbers. It may vary from specific numeric confidence regarding specific decision making paradigms. Very few studies directly ask participants to rate their confidence about their decisions and answers in response to specific scenarios. Therefore, this study intends to examine the relationship between objective numeracy and subjective confidence in four studies of Peters et al. (2006). We hypothesized that objective numeracy is positively associated with confidence in each study.

## Extension: Numeracy as a continuous measure

We added analyses to treat numeracy as a continuous variable instead of the dichotomy used in the target article. Methodologists have increasingly expressed concerns regarding the dichotomization of continous variables as it might result in suboptimal interpretations (Altman & Royston, 2006; Fedorov et al., 2009; Lazic, 2018; Mariooryad & Busso, 2015). One of the primary limitations is the loss of information, and treating samples within the same group as having the same underlying properties.

Peters et al. (2006) conducted a median split of numeracy scores: Participants who achieved a score of 9 or more were categorized as highly numerate whereas those who achieved 8 or lower were categorized as less numerate. However, the differences between individuals who achieved 8 and 9 might be neglectable, and no different than the differences between individuals who achieved 9 compared to 10 or 7 compared to 8. In addition, dichotomization reduces the power of statistical tests and effect sizes (Bakhshi et al., 2012). Fedorov et al. (2009) argued that 100 continuous observations are statistically equivalent to 158 dichotomized observations. Thus, the aim of treating numeracy as a continuous variable is to obtain more accurate effects, maximize power, and address potential misinterpretations resulting from dichotomization.

## Pre-registration and open-science

We pre-registered the experiment on the Open Science Framework (OSF) and data collection was launched later that week. Pre-registrations, power analyses, and all materials used in these experiments are available in the supplementary materials. We provided all materials, data, code, and pre-registration on: <https://osf.io/4hjck/> .

We provided additional open-science details and disclosures in the supplementary materials under “Open Science disclosures” sub-section. All measures, manipulations, exclusions conducted for this investigation are reported, all studies were pre-registered with power analyses, and data collection was completed before analyses.

# Method

**[IMPORTANT:   
Method and Results sections were written in the past tense using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection.]**

## Power analysis

We calculated effect sizes (ES) and power based on the statistics reported in the target article (see supplementary materials). We then conducted a power analysis using G\*Power (Faul, Erdfelder, Lang, & Buchner, 2007) for the statistical tests in each of the decision-making risk paradigms separately (i.e. framing effect, frequency-percentage effect, ratio bias and bets effect).

Power analyses were conducted on the results of the main findings in the original study that yielded significant effect and supported the hypotheses for Studies 1 to 4. The largest required sample size in all effects was a result from two-way between-subjective ANOVA, which when aiming for a power of 0.95 and alpha of 0.05 one-tail was *N* = 314. We provide further information regarding our calculations in the “Power analysis of original study effect to assess required sample for replication” section in the supplementary materials.

Given the possibility that the original effects are overestimated, we used the suggested Simonsohn (2015) rule of thumb, even if meant for other designs, and multiplied 314 by 2.5 resulting in 785 participants. Allowing for possible exclusions we summarized a total sample of 850 participants. Our sensitivity analysis indicated that a sample of 850 would allow the detection of *f* = 0.12 (one covariate, groups = 2, df = 1, 95% power, alpha = 5%, one-tail), an effect much weaker than any of the effects reported in the original, and the detection of *r* = 0.12 in our continuous measures extension, an effect considered weak in social psychology (Lovakov & Agadullina, 2021).

To demonstrate the results after data collection we simulated a dataset of 1000 participants using Qualtrics and report our analyses below based on that dataset. Results will later be updated to a sample of 850 and the real data.

## 

We will recruit participants from Amazon Mechanical Turk using the CloudResearch/Turkprime platform (Litman, Robinson, & Abberbock, 2017). Based on our extensive experience of running similar judgment and decision making replications on MTurk, to ensure high quality data collection, we will employ the following CloudResearch options: Duplicate IP Block. Duplicate Geocode Block, Suspicious Geocode Block, Verify Worker Country Location, Enhanced Privacy, CloudResearch Approved Participants and Block Low Quality Participants. We will also employ the [Qualtrics fraud and spam prevention measures](https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/): reCAPTCHA, prevent multiple submission, prevent ballotstuffing, bot detection, security scan monitor and relevantID.

Assignment pay is based on the federal wage of 7.25USD/hour, per minute, so for example - 5-8 minutes survey would be paid 1 USD per participant. We first pretested survey duration with 30 participants to make sure our time run estimate was accurate and adjusted pay as needed, the data of the 30 participants was not analyzed other than to assess survey completion duration and needed pay adjustments. [For those pretest participants, if survey duration was longer than expected, they would be paid a bonus as pay adjustment.]

Table 3

*Differences and similarities between original study and replication*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Peters et al. (2006)  study 1 | Peters et al. (2006)  study 2 and 3 | Peters et al. (2006)  study 4 | US MTurk workers |
| Sample size | 100 | 46 | 171 | 1000 | |
| Geographic origin | US American |  |  | US American | |
| Gender | 55 males, 45 females | Not reported | 79 males, 92 females | 256 males, 252 females, 492 other/did not disclose | |
| Median age (years) | Not reported | Not reported | Not reported | 48.00 | |
| Average age (years) | 26 | Not reported | 19 | 48.37 | |
| Standard deviation age (years) | Not reported | Not reported | Not reported | 28.96 | |
| Age range (years) | Not reported | Not reported | Not reported | 0-100 | |
| Medium (location) | Pencil and paper | Not reported | Not reported | Computer (online) | |
| Compensation | $10 | Not reported | Not reported | Nominal payment | |
| Year | 2005 | 2005 | 2005 | 2022 | |

n

## Design: Replication and Extension

We summarized the experimental design in Tables 4, 5, 6, and 7. To conduct a replication of the four studies in the original article, we will run the four studies together in a single data collection. The display of scenarios and conditions were counterbalanced using the randomizer “evenly present” function in Qualtrics. Scenarios were presented in random order and participants were randomly and evenly assigned into different conditions. This method was previously tested successfully in many of the replications and extensions conducted by our team (e.g., Adelina & Feldman, 2022; Vonasch et al., 2022; Yeung & Feldman, 2022), and is especially powerful in addressing concerns about the target sample (e.g., naivety and attentiveness) when some studies replicate successful whereas others do not, as well in drawing inferences about the links between the different studies and consistency in participants’ responding to similar decision-making paradigms.

[*For review: The Qualtrics survey .QSF file and an exported DOCX file are provided on the OSF folder. A preview link of the Qualtrics survey is provided on:* [*https://hku.au1.qualtrics.com/jfe/preview/SV\_0NaWp7LjCTgmO7I?Q\_CHL=preview&Q\_SurveyVersionID=current*](https://hku.au1.qualtrics.com/jfe/preview/SV_0NaWp7LjCTgmO7I?Q_CHL=preview&Q_SurveyVersionID=current)]

Table 4

*Study 1: Replication and extension experimental design*

|  |  |
| --- | --- |
| IV1: Numeracy  [between subject/continuous]  IV2: Positive-negative framing [between subject] | **IV1: Numeracy**  Original numeracy scale.  Extension numeracy scale. |
| **IV2: Positive framing condition**  Scores framed positively “% correct”  . | **Dependent variable**  Evaluation of students’ performance  Please rate each student’s quality of work  "*Very poor*" (-3) to "*Very good*" (3) (for each of the five students)  **Extension dependent variable**  Evaluation of subjective confidence level  How confident are you that you made an accurate assessment of the five students?  “*Not at all confident*” (0) to “*Very confident*” (6) |
| **IV2: Negative framing condition**  Scores framed negatively “% incorrect” |

Table 5

*Study 2: Replication and extension experimental design*

|  |  |
| --- | --- |
| IV1: Numeracy  [between subject/continuous]  IV2: Frequency-percentage description (risk format) [between subject] | **IV1: Numeracy**  Original numeracy scale.  Extension numeracy scale. |
| **IV2: Frequency condition**  “Of every 100… 10 are estimated…” | **Dependent variable**  Evaluation of risk level  Please rate the level of risk that Mr. Jones would harm someone “*Low risk*” (1) to “*High risk*” (6)  **Extension dependent variable**  Evaluation of subjective confidence level  How confident are you that made an accurate risk assessment? “*Not at all confident*” (0) to “*Very confident*” (6) |
| **IV2: Percentage condition**  “Of every 100… 10% are estimated… |

Table 6

*Study 3: Replication and extension experimental design*

|  |
| --- |
| **IV: Numeracy** [between subject/continuous]  Original numeracy scale.  Extension numeracy scale. |
| **Dependent variables**  Preference of bowl  Which bowl would you prefer to choose from?  “*Strong preference for Bowl A*” (-6) to “*Strong preference for Bowl B*” (6)  Affect precision for Bowl A choice  How clear a feeling do you have about the goodness or badness of Bowl A’s 9% chance of winning?  “*Completely unclear*” (0) to “*Completely clear*” (6)  Affect for Bowl A choice  How good or bad does Bowl A’s 9% chance of winning make you feel?  “*Very bad*” (-3) to “*Very good*” (3)  [**Added adjustment dependent variables**  Affect precision for Bowl B choice  How clear a feeling do you have about the goodness or badness of Bowl B’s 10% chance of winning?  “*Completely unclear*” (0) to “*Completely clear*” (6)  Affect for Bowl B choice  How good or bad does Bowl B’s 10% chance of winning make you feel?  “*Very bad*” (-3) to “*Very good*” (3) ]  **Extension dependent variables**  Evaluation of subjective confidence level  How confident are you that made an accurate risk assessment?  “*Not at all confident*” (0) to “*Very confident*” (6) |
|

Table 7

*Study 4: Replication and extension experimental design*

|  |  |
| --- | --- |
| IV1: Numeracy  [between subject/continuous]  IV2: Bet type (loss vs. no-loss) [between subject] | **IV1: Numeracy**  Original numeracy scale.  Extension numeracy scale. |
| **IV2: Bet - No loss condition**  “There is a 7/36 chance to win $9 and 29/36 chance to win nothing.” | **Dependent variable**  Evaluation of bet’s attractiveness  Please indicate your opinion of this bet’s attractiveness  "*Not at all attractive bet*" (0) to "*Extremely attractive bet*" (20)  Affect precision for bet  How clear a feeling do you have about the goodness or badness of the bet?  “*Completely unclear*” (0) to “*Completely clear*” (6)  Affect for bet  How good or bad does the bet make you feel?  “*Very bad*” (-3) to “*Very good*” (3)  **Extension dependent variable**  Evaluation of subjective confidence level  How confident are you that you made an accurate assessment of the five students?  “*Not at all confident*” (0) to “*Very confident*” (6) |
| **IV2: Bet - Loss condition**  “There is a 7/36 chance to win $9 and 29/36 chance to lose 5 cents.” |

## Procedures

Participants first read the consent form, study outline, and then acknowledged a warning about not looking up answers online. They were then randomly assigned to a condition in each of the four studies. The order of four studies and their conditions were randomized. After the completion of tasks of four scenarios, they completed two numeracy scales, in random order. Then, they verified not using external aids in answering the questionnaire. At the end, participants answered a number of funneling questions (seriousness towards the survey, study purpose conjecture, and feedback) and provided demographic information. We added a more comprehensive overview of the survey procedure in the “procedure” section in the supplementary.

## Measures

We detailed the measures of the replications and extensions for each condition in Tables 4, 5, 6, and 7. We provided all materials, with all experimental manipulation and the scales used, in the supplementary materials.

### Numeracy

Objective numeracy predictor was measured using the Numeracy Scale developed by Lipkus et al. (2001) (Cronbach’s = TBD). It has 11 items and the total mark is 11.

We added an additional numeracy measure as an extension: Numeracy Scale developed by Weller et al. (2013) (Cronbach’s = TBD). We refer to it as the “Rasch-based numeracy scale”, and it has eight items and the total mark is 8.

## Manipulations

### Study 1: Positive versus negative framing

Participants were randomly assigned to either positive framing or negative framing conditions. They were asked to rate the quality of five psychology students’ exam scores framed positively or negatively. The order of five exam scores was randomized.

### Study 2: Frequency and percentage condition

Participants were randomly assigned to frequency or percentage conditions. Participants read the scenario of Mr. Jones, a mental patient with the potential to harm someone when released. Participants then rate the risk level of patients like Mr. Jones under either frequency framing (i.e., 10 out of 100 patients) or percentage framing (i.e., 10% of 100 patients).

### Study 3: Ratio bias

Participants first read a scenario describing two jellybean bowls. Bowl A is the more attractive yet with less objectively favorable outcome than Bowl B. Participants rated their preference for Bowl A, and then chose one of the bowls. They then rated affect levels and affect precision of both bowls.

### Study 4: No loss versus loss condition

Participants were randomly assigned to loss versus no-loss conditions. Participants read the scenario on a bet with “a chance 7 out 36 chance to win $9 and 29 out 36 chance to win nothing” or with “a chance 7 out 36 chance to win $9 but 29 out 36 chance to lose 5 cents”. The chance of bet was visualized using a picture of a roulette wheel. Participants evaluated the attractiveness of the bet, and then rated affect and affect precision.

## Deviations

We note we made several adjustments that are deviations from the original’s design. We summarize the details of the deviations with comparisons of the original paper and our replication in Table 8.

In terms of the measurement of numeracy, we added an objective numeracy scale developed by Weller et al. (2013). The rationale for this extension is that this scale has demonstrated sound psychometric properties based on Rasch analysis and is argued to have better predictive validity than previous scales. Several recent studies have adopted it and shown support for high internal consistency (Cheng, 2020; Dolan et al., 2016; Peters et al., 2019).

We added a warning pledge block at the beginning of the questionnaire to ask participants not to look for answers and added a question at the end asking participants whether they used any external aids to search answers after the completion of two numeracy scales.

We made minor visual adjustments in the original numeracy scale (Lipkus et al., 2001), we removed decimals in 10.00, and turned the 1,000 into 1000. Given that we asked for and validated the input of numbers without decimals and commas, these may confuse participants.

The target article paper used SAT scores as a proxy measure for intelligence as they demonstrated that intelligence is positively associated with objective numeracy. Collection of SAT scores is not applicable to our target sample, and is not a core component of the target article.

The target article ran data collection for each of the studies separately using pencil and paper. We conducted data collection online in a unified design in which participants answer all the dependent variables of the four studies in random order.

In Study 1, the original paper did not report the specific scores of five psychology students, which only had one example. Therefore, we reconstructed our own version of the four students’ scores with four percent increment or four percent decrement (i.e., 66%, 70%, 74%, 78% and 82%).

In Study 3, we also added the questions of affect precision and affect for Bowl B. These were meant as exploratory measures to allow us to determine how participants feel about both options to allow for baseline comparisons. We considered the possibility that drawing conclusions from the ratings of only one of the two bowls may be lacking, whereas a comparison of the two options would be more accurate.

|  |  |  |
| --- | --- | --- |
| Hypothesis | Same+extension | We ran the original analyses and added a reframing of the hypotheses treating numeracy as a continuous variable. |
|  |  |  |
| DV construct | Similar | We reconstructed our version of the scores of the four students described in Study 1, as the stimuli was not provided in the article. |
| IV operationalization | Similar | We randomized the order of the numeracy questions |
| DV operationalization | Similar | In Study 3, we added exploratory extra questions for more optimal choice on affect and affect precision (on Bowl B). |
|  |  |  |
|  |  |  |
| Procedural details | Similar+extensions | The dependent variables on the four studies are completed together, in random order  Added a warning pledge before test  Added a question confirming not using external aids to find answers  Add familiarity questions in Studies 2, 3, and 4  We did not collect SAT scores |
| Physical settings | Different | Online questionnaire |
| Contextual variables | Different |  |
|  |  |  |
| Replication classification | Close replication |  |

We summarized the replication as a close replication using the criteria by LeBel et al., (2018) criteria, summarized in the supplementary materials in section “Replication closeness”.

## Data analysis strategy

### Replication: As in the original

The original paper dichotomized the numeracy scores as high numerate and low numerate. They conducted a 2 between x 5 within mixed ANOVA in Study 1, a two-way ANOVA in Study 2, and in Study 3 a chi-square test on the question of choosing Bowl, and an independent t-test to test bowl preference, affect, and affect precision. In Study 4, they conducted a factorial ANOVA for main interaction effect (i.e., numeracy and attractiveness of bet, numeracy and affect, numeracy and affect precision) and independent t tests to compare the responses (i.e., rate of attractiveness, affect and affect precision) of the high numerate under two conditions.

|  |  |  |
| --- | --- | --- |



|  |  |  |
| --- | --- | --- |



|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |



### Replication: Additional analyses

To our understanding, one of the major weaknesses of the target article is in their decision to dichotomize a continuous measure. In the replication, we supplemented the original analyses with additional analyses treating numeracy scores as intended - a continuous variable. Therefore, in Studies 1, 2, 3 and 4 we conducted correlational analyses and in Study 3 we conducted an extra independent t-test for bowl choice.

### Extensions

We conducted correlational analyses for confidence level and numeracy scale score.

### Exclusion criteria

We added questions regarding familiarity with the decision-making paradigms at the end of numeracy scales and each study. The whole responses of participants who answer “yes” on the numeracy scale in the original paper will be excluded. Participants who answer “yes” on the numeracy scale of extension or other studies and their responses will be only excluded correspondingly. Details exclusion criteria is provided in the “exclusion criteria” section in the supplementary.

# 

# Results

**[IMPORTANT:   
Method and Results sections were written in the past tense using a randomized dataset produced by Qualtrics and R codes to simulate what these sections will look like after data collection. These will be updated following the data collection.]**

Descriptive statistics of all measures are presented in Table 9. Statistical tests of the hypotheses are summarized in Table 10 11, 12, 13, 14 and 15. We did not apply the exclusion criteria for the simulated dataset and this will be completed after the data collection.

Table 9

*Studies 1-4: Descriptive statistics*

|  |  |  |  |
| --- | --- | --- | --- |
| Study | Conditions |  | Overall |
| 1 | Positive framing | Negative framing |  |
|  | *M =* 47.29*, SD =* 28.73*, N =* 500 | *M =* 49.44*, SD =* 29.19*, N =* 500 | *M =* 48.37*, SD =* 28.71*, N =* 1000 |
| 2 | Frequency condition | Percentage condition |  |
|  | *M =* 49.90*, SD =* 28.80*, N = 500* | *M =* 46.83*, SD =* 29.07*, N =* 500 | *M =* 48.37*, SD =* 28.71*, N =* 1000 |
| 3 | N/A | | *M = 48.37, SD =* 28.71*, N =* 1000 |
| 4 | Bet - No loss Condition | Bet - Loss condition |  |
|  | *M =* 49.18*, SD =* 29.22*, N =* 500 | *M =* 47.55*, SD =* 28.71*, N =* 500 | *M =* 48.37*, SD =* 28.71*, N =* 1000 |

*Note*. *N* = number of participants, *M* = mean, *SD* = standard deviation

## Replication: Original’s analyses with dichotomized numeracy

We first conducted statistical analyses following the methods used in the original article which dichotomized the continuous measure of numeracy into high numerate and low numerate via median split. Our Qualtrics simulated dataset failed to generate a sufficient range of numeracy scores, and we therefore created a new variable where we randomized a numeracy score from 0 to 11 using R and used this variable in the following data analysis. The median of numeracy scores was 5 (mean = 5.5, range = 0-11). Therefore, participants were the highly numerate whose overall score above 5 and were low numerate whose overall score equal to or below 5.

In Study 1, we conducted a mixed ANOVA and found no support for a main effect of numeracy on framing effect (*F*(1, 996) = 0.02, *p* = .911, *η²* = 0.00, 90% CI [0.00, 0.00]). We failed to find support for the hypothesis that the less numerate show a stronger framing effect than the highly numerate.

In Study 2, we conducted a factorial ANOVA and found no support for a main effect of numeracy on frequency-percentage effect (*F*(1, 996) = 1.05, *p* = .979, *η²* = 0.00, 90% CI [0.00, 0.01]). We failed to find support for the hypothesis that the less numerate are affected more by the frequency-percentage effect than the highly numerate.

In Study 3, we conducted a Chi-squared test of independence and found no support for the association between numeracy and Bowl choice (*χ2*(1), *p* = .331, *φ* = 0.03, 95% CI [-0.03, 0.09] ). We failed to find support for the hypothesis that the less numerate make less optimal choices in competing affective decisions than the highly numerate. In addition, we conducted an independent samples t-test and found no support for differences in attractiveness of the bet between the highly numerate and the low numerate (*t*(998) = 0.11, *p* = .912, *d* = 0.01, 95% CI [-0.48, 0.53]). We also found no support for differences in affect precision for Bowl A (*t*(998) = -0.29, *p* = .769, *d* = -0.02, 95% CI [-0.29, 0.21]). Therefore, we failed to find support for the hypothesis that the less numerate makes lower affective precision in competing affective decisions than the highly numerate.

In Study 4, we conducted the factorial ANOVA and found no support for main effects of numeracy on attractiveness, affect precision, and affect of no-loss and loss bets (*F*(1, 996) = 0.92, *p* = .337, *η²* = 0.00, 90% CI [0.00, 0.01]; *F*(1, 996) = 0.10, *p* = .751, *η²* = 0.00, 90% CI [0.00, 0.00]); *F*(1, 996) = 0.28, *p* = .599, *η²* = 0.00, 90% CI [0.00, 0.00]). Therefore, we failed to find support for the hypothesis that the less numerate makes more optimal choices, shows lower affective precision, and draws less affect in probabilities and numerical comparisons than the highly numerate.

We summarized the results of Studies 1, 2, and 4 in Table 10 and the results of Study 3 in Table 11.

*Studies 1, 2 and 4: Summary of statistical tests*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *F* | *df* | *p* | *η²* and CI | Interpretation |
| Study 1 (Mixed ANOVA) | | | | | |
| Numeracy (High/Low) and framing effect (Positive/Negative framing condition) | 0.02 | 1, 996 | .911 | 0.00  [0.00, 0.00] |  |
| Study 2 (Factorial ANOVA) | | | | | |
| Numeracy (High/Low) and frequency-percentage effect (Frequency/Percentage condition) | 1.05 | 1, 996 | .979 | 0.00  [0.00, 0.01] |  |
| Study 4 (Factorial ANOVA) | | | | | |
| Numeracy (High/Low) and attractiveness of bet (No loss/Loss condition) | 0.92 | 1, 996 | .337 | 0.00  [0.00, 0.01] |  |
| Numeracy (High/Low) and affect precision of bet (No loss/Loss condition) | 0.10 | 1, 996 | .751 | 0.00  [0.00, 0.00] |  |
| Numeracy (High/Low) and affect of bet (No loss/Loss condition) | 0.28 | 1, 996 | .599 | 0.00  [0.00, 0.00] |  |

CI = 90% confidence intervals

Table 11

*Study 3: Summary of statistical tests*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Chi-squared test | *χ2* | *df* | *p* | *φ* and CI | Interpretation |
| Bowl Choice | 0.95 | 1 | .331 | 0.03 [-0.03, 0.09] |  |
| Independent t test | *t* | *df* | *p* | *d* and CI | Interpretation |
| Numeracy (High/Low) and preference | 0.11 | 998 | .912 | 0.01 [-0.48, 0.53] |  |
| Numeracy (High/Low) and affect precision for Bowl A | -0.29 | 998 | .769 | -0.02 [-0.29, 0.21] |  |
| Numeracy (High/Low) and affect for Bowl A | 1.28 | 998 | .202 | 0.08 [-0.09, 0.40] |  |

## Replication: Analyses using continuous numeracy

### Original numeracy scale

In Study 1, we failed to find support for associations between numeracy and ratings of students in both the positive and the negative framing conditions (*r* = -0.03, 95% CI [-0.12, 0.05], *p* = .445; *r* = -0.02, 95% CI [-0.11, 0.06], *p* = .598). We then compared two correlations with the tool “cocor” (Diedenhofen, & Musch, 2015) and found no support for any differences (*z* = -0.16, *p* = .875). Therefore, we concluded we failed to find support for the relationship between numeracy and framing effect.

In Study 2, we found no support for associations between numeracy and ratings of risk level in both frequency and percentage conditions (*r* = -0.06, 95% CI [-0.14, 0.03], *p* = .219; *r* = 0.02, 95% CI [-0.07, 0.10], *p* = .751). We found no support for differences between the correlations (*z* = -0.26, *p* = .207). Therefore, we concluded we failed to show the relationship between numeracy and frequency-percentage effect.

In Study 3, we conducted an independent t-test between numeracy and bowl choice and found no difference between choices of Bowl A and Bowl B (*t*(998) = -0.16, *p* = .875, *d* = 0.05, 95% CI [-0.07, 0.18]). We then found no support for associations between numeracy and bowl preference (*r* = 0.00, 95% CI [-0.06, 0.06], *p* = .957) and affect precision for Bowl A (*r* = -0.01, 95% CI [-0.07, 0.06], *p* = .841).



In Study 4, we found no support for associations between numeracy and attractiveness of bet, affect precision, and affect for both the no loss and loss conditions. We found no support for differences between the correlations with attractiveness, affect precision, and affect (*z* = 0.32, *p* = .752; *z* = 0.63, *p* = .528; *z* = -1.58, *p* = .115).



### 

*[The results using rasch-based numeracy scale were the same as using the original numeracy scale. We will report relevant results after data collection. They were shown in Table 12, 13 and 14 as well. ]*

Table 12

*Studies 1, 2, 3, and 4: Summary of statistical tests*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | *r* and CI | | *p* | Spearman’s rho | Interpretation |
| Original numeracy scale | |  |  |  |  | |  |
| Study 1 | Rating of students in Positive framing condition | | -0.03  [-0.12, 0.05] | | .445 | -0.03 |  |
|  | Rating of students in Negative framing condition | | -0.02  [-0.11, 0.06] | | .598 | -0.02 |  |
| Study 2 | Risk rating in  Frequency condition | | -0.06  [-0.14, 0.03] | | .219 | -0.06 |  |
|  | Risk rating in Percentage condition | | 0.02  [-0.07, 0.10] | | .751 | 0.01 |  |
| Study 3 | Bowl preference | | 0.00  [-0.06, 0.06] | | .957 | 0.00 |  |
|  | Affect precision for Bowl A | | 0.01  [-0.07, 0.05] | | .680 | 0.01 |  |
|  | Affect for Bowl A | | -0.02  [-0.08, 0.04] | | .483 | -0.02 |  |
| Study 4 | No Loss condition | | | | | | |
|  | Attractiveness | | 0.04  [-0.05, 0.13] | | .399 | 0.04 |  |
|  | Affect precision | | 0.04  [-0.05, 0.13] | | .383 | 0.04 |  |
|  | Affect | | -0.07  [-0.16, 0.02] | | .123 | -0.07 |  |
|  | Loss condition | |  | |  |  |  |
|  | Attractiveness | | 0.02  [-0.06, 0.11] | | .592 | 0.03 |  |
|  | Affect precision | | 0.00  [-0.08, 0.09] | | .944 | 0.00 |  |
|  | Affect | | 0.03 [-0.06, 0.12] | | .512 | 0.03 |  |
| Rasch-based numeracy scale | | | | | | | |
| Study 1 | Rating of students in Positive framing condition | | 0.02  [-0.07, 0.11] | | .653 | 0.02 |  |
|  | Rating of students in Negative framing condition | | 0.01  [-0.08, 0.10] | | .862 | 0.02 |  |
| Study 2 | Risk rating in  Frequency condition | | -0.05  [-0.14, 0.04] | | .263 | -0.05 |  |
|  | Risk rating in Percentage condition | | 0.01  [-0.08, 0.09] | | .887 | 0.01 |  |
| Study 3 | Bowl preference | | 0.00  [-0.06, 0.06] | | .957 | 0.00 |  |
|  | Affect precision for Bowl A | | -0.01  [-0.07, 0.06] | | .841 | -0.01 |  |
|  | Affect for Bowl A | | -0.06  [-0.12, 0.00] | | .068 | -0.06 |  |
| Study 4 | No Loss condition | | | | | | |
|  | Attractiveness | | 0.00  [-0.08, 0.09] | | .936 | 0.00 |  |
|  | Affect precision | | 0.00  [-0.09, 0.09] | | .975 | 0.00 |  |
|  | Affect | | 0.01  [-0.08, 0.10] | | .787 | 0.01 |  |
|  | Loss condition | |  | |  |  |  |
|  | Attractiveness | | -0.04  [-0.13, 0.05] | | .388 | 0.03 |  |
|  | Affect precision | | 0.00  [-0.08, 0.09] | | .950 | 0.00 |  |
|  | Affect | | -0.02  [-0.11, 0.07] | | .612 | -0.02 |  |

*Note*. CI = 95% confidence intervals. Interpretation is using the LeBel et al. (2019) criteria.

Table 13

*Study 3: Optimal bowl choice*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Independent t-test | *t* | *df* | *p* | *d* and CI | Interpretation |
| Original numeracy scale | | | | |  |
| Bowl Choice | -0.16 | 998 | .875 | 0.05 [-0.07, 0.18] |  |
| Rasch-based numeracy scale | | | | | |
| Bowl Choice | 0.23 | 998 | .819 | 0.01 [-0.11, 0.14] |  |

*Note*

Table 14

*Studies 1, 2, and 4: Comparison of correlations*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Fisher’s *z* | *p* | Interpretation |
| Original numeracy scale | | | | |
| Study 1 | Framing effect | -0.16 | .875 |  |
| Study 2 | Frequency-percentage effect | -1.26 | .207 |  |
| Study 4 | Attractiveness | 0.32 | .752 |  |
|  | Affect precision | 0.63 | .528 |  |
|  | Affect | -1.58 | .115 |  |
| Rasch-based numeracy scale | | | | |
| Study 1 | Framing effect | 0.16 | .875 |  |
| Study 2 | Frequency-percentage effect | -0.95 | .344 |  |
| Study 4 | Attractiveness | 0.63 | .528 |  |
|  | Affect precision | 0.00 | 1.000 |  |
|  | Affect | 0.47 | .636 |  |

## Extension: Confidence

We examined the extension of the association between objective numeracy and subjective confidence with both original numeracy scale and rasch-based numeracy scale.

We conducted the correlational study but failed to find any associations between objective numeracy and confidence in all four studies (see Table 15 for detailed results).

Table 15

*Confidence: Summary of statistical tests in Studies 1-4*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Correlation |  | *r* and CI | *p* | Spearman’s rho | Interpretation |
| Original numeracy scale and Confidence level | | | | | |
| Study 1 | Positive framing condition | 0.03  [-0.05, 0.12] | .436 | 0.03 |  |
|  | Negative framing condition | 0.06  [-0.02, 0.15] | .158 | 0.06 |  |
| Study 2 | Frequency condition | 0.06  [-0.02, 0.15] | .150 | 0.06 |  |
|  | Percentage condition | -0.04  [-0.13, 0.05] | .341 | -0.04 |  |
| Study 3 |  | 0.00  [-0.06, 0.07] | .912 | 0.01 |  |
| Study 4 | No loss condition | -0.01  [-0.10, 0.08] | .836 | -0.01 |  |
|  | Loss condition | 0.03  [-0.06, 0.12] | .479 | 0.03 |  |
| Rasch-based numeracy scale and Confidence level | | | | | |
| Study 1 | Positive framing condition | 0.03  [-0.05, 0.12] | .436 | 0.03 |  |
|  | Negative framing condition | 0.06  [-0.02, 0.15] | .158 | 0.150 |  |
| Study 2 | Frequency condition | 0.06  [-0.02, 0.15] | .150 | 0.06 |  |
|  | Percentage condition | -0.04  [-0.13, 0.05] | .341 | -0.04 |  |
| Study 3 |  | 0.00  [-0.06, 0.07] | .912 | 0.01 |  |
| Study 4 | No loss condition | 0.04  [-0.04, 0.13] | .325 | 0.04 |  |
|  | Loss condition | -0.03  [-0.12, 0.06] | .506 | -0.03 |  |

*Note*. CI = 95% confidence intervals

## Assumptions Checks

We will use the Levene’s test to check the homogeneity of variances and the Shapiro-Wilks test to check the normality of variables for ANOVA and independent t-test. We will do the appropriate data processing (e.g. data transformation) according to how assumptions are violated. This part will be completed after data collection. In addition, we supplemented the analyses with a report of Spearman correlations, shown in previous tables already, if the assumptions of correlation are violated. We will add results of assumptions check results in the Assumption check section in the supplementary.



## Exploratory analyses

The exploratory analysis will be completed in Stage 2 after data collection.

We will examine order effects if we fail to find support for any of the studies. We will run an analysis with display order as a covariate, to try and see whether the order impacted the effect size.

We added extra questions for affective arousal and affect precision for Bowl B in Study 3. We found no support for differences for both affect precision and affect towards Bowl A and Bowl B (*t*(999) = -1.60, *p* = .110, *d* = -0.05, 95% CI [-0.32, 0.03]; *t*(999) = 0.67, *p* = .506, *d* = 0.02, 95% CI [-0.11, 0.23]).

We will compare the original effect (e.g. *d*) and extension effect (e.g. *r*) with an approximation conversion. We will report both pre and post conversion effect sizes.

We also added timers in two numeracy measures and will conduct the exploratory reaction time analyses.

## Comparing replication to original findings

*[Since the simulated dataset barely generated any meaningful results, the comparison between replication and original findings will be completed after data collection. We will describe whether the replication successfully replicated the original findings as well as compare the results of the replication to original findings for different hypotheses.]*

# Discussion

**[IMPORTANT:   
Discussion section is only to be completed in Stage 2 following data collection]**

We conducted a pre-registered replication of numeracy. The results are [consistent/not consistent/partially consistent and partially inconsistent] with the original results .

## Replication

Overall, we found that … . In summary, the goal of the project was to assess the replicability of the research presented by Peters et al. (2006) in support of numeracy.

## 

We ran extensions examining the relationship between objective numeracy and numeric confidence under specific conditions. Overall, our findings showed that XXX. We found [weak to no/weak/medium/strong] support for our hypothesis.

## Implications, limitations, and directions for future research

[In this section, we will discuss the implications, limitations and future directions of our research. We will discuss, but not limited to, several methodological issues concerning our replications. For example, we could not ensure that participants use online shortcuts or calculator to answer numerical questions. In addition, we will also discuss the potential weakness of individual differences or between-participants variability. ]

# References

Adelina, N., & Feldman, G. (2021). Are past and future selves perceived differently from present self? Replication and extension of Pronin and Ross (2006) temporal differences in trait self-ascriptions. *International Review of Social Psychology*, 34(1): 29, 1–16. DOI: 10.5334/irsp.571

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *Bmj*, *332*(7549), 1080. https://doi.org/10.1136/bmj.332.7549.1080

Bakhshi, E., McArdle, B., Mohammad, K., Seifi, B., & Biglarian, A. (2012). Let continuous outcome variables remain continuous. *Computational and mathematical methods in medicine*, *2012*. https://doi.org/10.1155/2012/639124

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... & Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication?. Journal of Experimental Social Psychology, 50, 217-224.

Cheng, J. (2020). The role of numeracy and impulsivity in intertemporal choice and decision making. *Psychological Thought, 13*(1), 254-272. https://doi.org/10.37708/psyct.v13i1.442

Choi, H., Wong, J. B., Mendiratta, A., Heiman, G. A., & Hamberger, M. J. (2011). Numeracy and framing bias in epilepsy. *Epilepsy & Behavior, 20*(1), 29– 33. doi: 10.1016/j.yebeh.2010.10.005

Dickert, S., Kleber, J., Peters, E., & Slovic, P. (2011). Numeracy as a precursor to pro- social behavior: The impact of numeracy and presentation format on the cognitive mechanisms underlying donation decisions. *Judgment and Decision Making, 6*(7), 638– 650.

Diedenhofen, B. & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. PLoS ONE, 10(4): e0121945. https://doi.org/10.1371/journal.pone.0131499

Dolan, J. G., Cherkasky, O. A., Li, Q., Chin, N., & Veazie, P. J. (2016). Should health numeracy be assessed objectively or subjectively?. *Medical Decision Making, 36*(7), 868-875. doi: 10.1177/0272989X15584332

Estrada-Mejía, de Vries, & Zeelenberg Estrada-Mejía C., de Vries M., & Zeelenberg M. (2016). Numeracy and wealth. *Journal of Economic Psychology, 54*, 53–63. https://doi.org/10.1016/j.joep.2016.02.011

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

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods, 39(2), 175-191. <https://doi.org/10.3758/BF03193146>

Fedorov, V., Mannino, F., & Zhang, R. (2009). Consequences of dichotomization. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry, 8*(1), 50-61. https://doi.org/10.1002/pst.331

Freling, T. H., Vincent, L. H., & Henard, D. H. (2014). When not to accentuate the positive: Re-examining valence effects in attribute framing. *Organizational Behavior and Human Decision Processes, 124*(2), 95-109. https://doi.org/10.1016/j.obhdp.2013.12.007

Gamliel, E., & Kreiner, H. (2017). Outcome proportions, numeracy, and attribute- framing bias. *Australian Journal of Psychology, 69*(4), 283– 292. https://doi.org/10.1111/ajpy.12151

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– 116. doi: 10.1002/ijop.12138

Garcia-Retamero, R., Andrade, A., Sharit, J., & Ruiz, J. G. (2015). Is patients’ numeracy related to physical and mental health?. Medical Decision Making, 35(4), 501-511. https://doi.org/10.1177/0272989X15578126

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases”. *European Review of Social Psychology, 2*, 83–115. https://doi.org/10.1080/14792779143000033

Hill, W. T., & Brase, G. L. (2012). When and for whom do frequencies facilitate performance? On the role of numerical literacy. *Quarterly journal of experimental psychology, 65*(12), 2343-2368. doi: 10.1080/17470218.2012.687004

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. American psychologist, 58(9), 697. <https://doi.org/10.1037/0003-066X.58.9.697>

Kleber, J., Dickert, S., Peters, E., & Florack, A. (2013). Same numbers, different meanings: How numeracy influences the importance of numbers for prosocial behavior. Journal of Experimental Social Psychology, 49, 699 –705. doi:10.1016/j.jesp.2013.02.009

Lazic, S. E. (2018). Four simple ways to increase power without increasing the sample size. *Laboratory animals*, *52*(6), 621-629. https://doi.org/10.1177/0023677218767478

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*, 389-402. https://doi.org/10.1177/2515245918787489

LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate replications. *Meta-Psychology*, 3, 1-9. doi: 10.15626/MP.2018.843

Levin, I. P. , & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15*(3), 374–378. https://doi.org/10.1086/20917

Levin, I. P. , Schneider, S. L. , & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human* *Decision Processes, 76*(2), 149–188. https://doi.org/10.1006/obhd.1998.2804

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making,* *25*, 361–381. doi:10.1002/bdm.752

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, 51(3), 485-504.

Mariooryad, S., & Busso, C. (2015). The cost of dichotomizing continuous labels for binary classification problems: Deriving a Bayesian-optimal classifier. *IEEE Transactions on Affective Computing*, *8*(1), 119-130. https://doi.org/10.1109/TAFFC.2015.2508454

Miller, D. T., Turnbull, W., & McFarland, C. (1989). When a coincidence is suspicious: The role of mental simulation. *Journal of Personality and Social Psychology, 57*, 581–589. https://doi.org/10.1037/0022-3514.57.4.581

Nelson, W. L., Moser, R. P., & Han, P. K. (2013). Exploring objective and subjective numeracy at a population level: findings from the 2007 Health Information National Trends Survey (HINTS). *Journal of Health Communication, 18*(2), 192-205. doi: 10.1080/10810730.2012.688450

Okamoto, M., Kyutoku, Y., Sawada, M., Clowney, L., Watanabe, E., Dan, I., & Kawamoto, K. (2012). Health numeracy in Japan: Measures of basic numeracy account for framing bias in a highly numerate population. *BMC Medical Informatics and Decision Making, 1*2(1), 104. https://doi.org/10.1186/1472-6947-12-104

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science, 21*(1), 31-35. https://doi.org/10.1177/0963721411429960

Peters, E., & Bjalkebring, P. (2015). Multiple numeric competencies: When a number is not just a number. Journal of personality and social psychology, 108(5), 802. doi: 10.1037/pspp0000019

Peters, E. (2020). *Innumeracy in the wild: Misunderstanding and misusing numbers.* Oxford University Press.

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

Peters, E., Hart, P. S., & Fraenkel, L. (2011). Informing patients: The influence of numeracy, framing, and format of side effect information on risk perceptions. *Medical Decision Making, 31*(3), 432– 436. doi: 10.1177/0272989X10391672

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

Peters, E., Tompkins, M. K., Knoll, M. A., 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, 116(39), 19386-19391. https://doi.org/10.1073/pnas.1903126116

Piñon, A., & Gambara, H. (2005). A meta-analytic review of framing effect: risky, attribute and goal framing. *Psicothema, 17*(2), 325-331.

Reyna, V. F., and Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences 18*, 89–107. doi: 10.1016/j.lindif.2007.03.011

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*(6), 943–973. https://doi.org/10.1037/a0017327

Rolison, J. J., Morsanyi, K., & Peters, E. (2020). Understanding health risk comprehension: The role of math anxiety, subjective numeracy, and objective numeracy. *Medical Decision Making, 40*(2), 222-234. https://doi.org/10.1177/0272989X20904725

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin, 119*(1), 3. https://doi.org/10.1037/0033-2909.119.1.3

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. Psychological science, 26(5), 559-569. https://doi.org/10.1177/0956797614567341

Traczyk, J., & Fulawka, K. (2016). Numeracy moderates the influence of task-irrelevant affect on probability weighting. *Cognition, 151*, 37-41. https://doi.org/10.1016/j.cognition.2016.03.002

Traczyk J., Sobkow A., Fulawka K., Kus J., Petrova D., & García-Retamero R. (2018). Numerate decision makers don’t use more effortful strategies unless it pays: A process tracing investigation of skilled and adaptive strategy selection in risky decision making. *Judgment and Decision Making, 13*(4), 372–381.

Tversky, A., & Kahneman, D. (1985). The Framing of Decisions and the Psychology of Choice. *Behavioral Decision Making*, 25-41. doi:10.1007/978-1-4613-2391-4\_2

Van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of experimental social psychology*, 67, 2-12.

Vonasch, A., Hung, W., Leung, W., Nguyen, T., Chan, S., Cheng, B., & Feldman‎, G. (2022). "Less is better" in separate evaluations versus "More is better" in joint evaluations: Mostly successful close replication and extension of Hsee (1998). DOI 10.17605/OSF.IO/9UWNS, retrieved from https://osf.io/nhyp9/

Woloshin, S., Schwartz, L. M., & Welch, H. G. (2005). Patients and medical statistics. *Journal of General Internal Medicine, 20*(11), 996-1000. doi: 10.1111/j.1525-1497.2005.00179.x

Yeung, S. & Feldman, G. (2022). Revisiting the Temporal Pattern of Regret: Replication of Gilovich and Medvec (1994) with extensions examining responsibility. DOI 10.17605/OSF.IO/7M3Q2, retrieved from https://osf.io/vncy7/

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41.