Judgment and Decision Making (2023), **Vol. 18**:e39 1–25
doi:10.1017/jdm.2023.40

REGISTERED REPORT

Working memory capacity and the risky-choice framing effect: A preregistered replication and extension of Corbin et al. (2010)

Boris Bogdanov[1], Jonathan Corbin[2], Sabina Dobreva[1], Todd McElroy[3] and Nikolay R. Rachev[1]

[1]Department of General, Experimental, Developmental, and Health Psychology, Sofia University St. Kliment Ohridski, Sofia, Bulgaria; [2]Center for Advanced Hindsight, Duke University, Durham, NC, USA and [3]Psychology Department and The Water School, Florida Gulf Coast University, Fort Myers, FL, USA

Corresponding author: Nikolay R. Rachev; Email: nrrachev@phls.uni-sofia.bg

Received: 11 October 2023; **Revised:** 20 October 2023; **Accepted:** 20 October 2023

Keywords: risky-choice framing; working memory capacity; replication; dual-process theories; fuzzy-trace theory; pragmatic inference account

Abstract
While working memory capacity is associated with superior performance on a number of tasks, could it paradoxically sometimes be associated with suboptimal performance? Corbin et al. (2010, *Judgment and Decision Making* 5(2), 110–115) found that, in a between-subjects design, higher WMC is associated with a larger risky-choice framing effect, traditionally conceived of as a departure from rational principles. Such surprising findings are of potentially great theoretical importance; yet the original study was underpowered. In this registered report, we aimed to replicate and extend the original findings, by conducting an online experiment among 425 North Americans. To extend the findings beyond the specific single tasks used in the original study, we used three WMC tasks with different processing components and six framing problems involving human lives. In a close replication, the frame significantly interacted with neither the Ospan short absolute score nor the Ospan short partial score in predicting ratings on the disease-framing problem. Similarly, in an extended replication, a composite WMC score did not significantly interact with the frame in predicting ratings on three framing problems involving human lives. The Bayes factors showed that the data were 3 to 10 times more likely under the null hypothesis of no interaction between WMC and frame. Taken together, these findings show an absence of association between the between-subjects risky-choice framing effect and WMC. This outcome is compatible with four out of the six theoretical accounts we considered, and is uniquely predicted by the default-interventionist dual-process account and the pragmatic inference account. Further research can more rigorously pit conflicting predictions of these accounts against each other.

1. Introduction

When people deal with risk, their choices are sometimes influenced by apparently irrelevant factors, such as the way the choice options are presented. In particular, people tend to choose a safe option when the information is presented as potential gains, but they prefer a risky option when the same information is presented as potential losses—a phenomenon known as the risky-choice framing effect (Levin et al., 1998; Tversky & Kahneman, 1981, see the Appendix for examples). The effect is well-established (Klein et al., 2014; Kühberger, 1998) and is typically treated as a deviation from rational

© The Author(s), 2023. Published by Cambridge University Press on behalf of Society for Judgment and Decision Making and European Association of Decision Making. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

principles (e.g., Tversky & Kahneman, 1986; but see Mandel, 2014). Although considerable research has been conducted on the individual differences in the effect (e.g., Lauriola & Levin, 2001; Levin et al., 2002; Simon et al., 2004), it has sometimes reported contradictory results (e.g., regarding the effect of cognitive style, Smith & Levin, 1996 vs. LeBoeuf & Shafir, 2003 and Levin et al., 2002; or regarding gender effects, Fagley & Miller, 1997 vs. Reyna et al., 2011). Other times, the available evidence is based on single studies using relatively small samples. For instance, Corbin et al. (2010) reported a larger framing effect among participants with higher versus lower working memory capacity (WMC), as measured by the automated operation span task (Ospan, Unsworth et al., 2005). This finding is surprising because most relevant research either found that cognitive capacity is positively associated with adhering to rational principles or failed to find any association. The finding is also inconsistent with a dominant account of the framing effect, namely the default-interventionist dual-process account (Kahneman, 2000, 2003), though it could be accounted for by alternative accounts such as the fuzzy-trace theory (Corbin et al., 2015; Reyna, 2012). We replicated and extended Corbin et al.'s (2010) study to provide more robust evidence, and thus better inform theoretical accounts about the relationship between WMC and risky-choice framing.

1.1. Empirical evidence

Working memory refers to the set of processes that hold a limited amount of information in a highly available state to be used in thoughts and actions (Cowan, 2017; Oberauer et al., 2018). Working memory plays a central role in cognitive processing, in particular in deliberate cognition (Oberauer et al., 2018). Although there are many definitions of working memory, and relatedly various experimental paradigms to study it (Cowan, 2017), some individuals perform consistently better than others on various working memory tasks. These individuals are said to have higher WMC. Performance on WMC measures is also strongly associated with measures of general intelligence (Oberauer et al., 2018), to the point that researchers of individual differences in judgment and decision-making often put them into the same category of 'cognitive ability/capacity' (e.g., Corbin et al., 2015; Evans & Stanovich, 2013; Stanovich & West, 2000).

Higher cognitive capacity is associated with better performance on a host of tasks related to probability judgment and decision-making under risk (Brevers et al., 2012; Cokely & Kelley, 2009; Corbin et al., 2015; Del Missier et al., 2013; Dougherty & Hunter, 2003; Starcke et al., 2011). Recently, Burgoyne et al. (2023) found that a latent rationality factor, composed of base rate, conjunction fallacy, and Wason selection tasks, was strongly associated with intelligence ($r = 0.56$) and WMC ($r = 0.44$). Yet, some judgment and decision-making tasks failed to display associations with cognitive capacity (see Stanovich & West, 2008, p. 686, for an illustrative summary of both positive and null findings). In a related series of studies, McElroy et al. (2020) found that more thought was associated with improved decision-making for complex tasks but not for simple ones, including risky-choice tasks.

Similarly, research on the risky-choice framing effect in particular has either found a negative association with cognitive capacity or failed to find any association. Whenever a negative association was found, it was always in within-subjects designs, where participants see both versions of the same problem at different points in time (Bruine de Bruin et al., 2007). Even when using within-subjects designs, research has sometimes reported mixed findings (Del Missier et al., 2012; Stanovich & West, 1998) or failed to find a significant association altogether (Toplak et al., 2014). Research varying the frame between subjects (as originally done by Tversky & Kahneman, 1981; and also by Corbin et al., 2010) has failed to find reliable differences in framing effects related to cognitive capacity (Stanovich & West, 2008). The framing effect was also not larger under memory load compared with no load (Whitney et al., 2008), suggesting that working memory considerations play little role in framing tasks at least in some contexts.

In contrast to negative or null findings, results like Corbin et al.'s (2010), pointing to a positive association between cognitive capacity and the framing effect, are rare. Stanovich and West (2008) found a descriptively larger framing effect among individuals with higher scores on the Scholastic

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

Assessment Test (SAT), as an index of cognitive ability. However, this finding was not statistically significant and was reported as a failure to find an association. Using a monetary task, Urs et al. (2019) failed to find a main effect of frame, and yet found that high-WMC individuals were more risk-seeking than low-WMC individuals under the loss frame when there was a large value at stake. The authors interpreted this finding as indicating a larger bias among high-WMC individuals, partially in accordance with Corbin et al. (2010). However, it is also possible that high-WMC individuals made superior choices because the risky options were also of a higher expected value than the sure options.

The strongest evidence supporting Corbin et al.'s (2010) findings comes from a further study by Corbin (2013, Experiment 1) that successfully replicated, using a larger sample (N = 161), the larger framing effect among higher-WMC individuals. However, the effect was present only among individuals scoring low in numeracy. Furthermore, the effect was not significant when using a score combining participants' choice (Program A vs. B) and their confidence. Under high memory demands, higher-WMC individuals also displayed a smaller rather than larger framing effect compared with lower-WMC individuals (Corbin, 2013, Experiment 2). Rather than unanimously confirming Corbin et al.'s (2010) findings, Corbin's (2013) findings present a mixed picture of possible moderators affecting the relationship between framing and WMC in either direction.

In sum, Corbin et al.'s (2010) findings of a positive association between cognitive capacity and the framing effect are corroborated by few studies, whereas most research has either shown a negative association or failed to find any association. Yet, if a positive relationship exists, the theoretical implications for advancing our understanding of the framing effect are of great importance.

1.2. Competing theoretical predictions

1.2.1. Dual-process theories

The framing effect was originally described and accounted for in the context of prospect theory (Tversky & Kahneman, 1981, 1986). Prospect theory proposes that people tend to evaluate the options relative to a neutral reference point, which shifts depending on how the options are presented. In the widely used disease problem (Tversky & Kahneman, 1981, see the Appendix), reading about lives to be saved induces a reference point of zero lives saved, so the options are evaluated as potential gains. By contrast, reading about people who would die induces a reference point of zero people dying, so the options are evaluated as potential losses. In addition, following a basic psychophysical law, people are decreasingly sensitive to changes of magnitude as the magnitude itself increases, which leads to risk aversion in the domain of gains and risk seeking in the domain of losses (Tversky & Kahneman, 1981). Dependence on reference points and diminished sensitivity jointly produce the risky-choice framing effect.

The prospect theory's interpretation of the framing effect was incorporated into the dual-process framework of thinking (Kahneman, 2011; Stanovich & West, 2000), distinguishing between an automatic and effortless System 1/Type 1 thinking and a conscious and effortful System 2/Type 2 thinking that is, crucially, dependent on the limited cognitive capacity (Evans & Stanovich, 2013). In particular, Kahneman (2011) regarded both reference point and diminishing sensitivity as features of System 1. Kahneman's (2003, 2011) account of the framing effect adheres to the so-called default-interventionist subtype of dual-process theories (Evans & Stanovich, 2013), which propose that an initial, default response to thinking problems is always provided by System 1. System 2 then intervenes, but only if it detects a conflict between the System 1 intuition and some normative consideration (Kahneman, 2000). Consequently, individuals who are better able and willing to think would do better on thinking tasks only if they are provided with a cue pointing to such conflict. When framing problems are presented between subjects, no cue is provided to realize that an alternative framing of the same problem exists (Kahneman, 2000, 2003). Both cognitively sophisticated and less sophisticated individuals would thus be equally likely to fall prey to the effect. By contrast, within-subjects designs help establish the equivalence between the two frames, and thus, provide an advantage to individuals of higher cognitive capacity, who are more willing or able to recognize the equivalence, recall their initial

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

response, and be consistent in their responses. Thus, depending on design, the default-interventionist view accounts for the different patterns of associations between cognitive capacity and the framing effect (see also Stanovich & West, 2008).

Other dual-process accounts might also be relevant to the association between the framing effect and cognitive capacity. For instance, the parallel-competitive approach (Sloman, 1996) proposes that Type 1 and Type 2 processes operate in parallel and provide competing responses to a problem. If Type 2 processes are involved from the start, it is possible that at least some participants spontaneously consider different frames, resulting in a diminished framing effect. Since Type 2 processes depend on cognitive capacity, those of higher capacity should be more likely to resist framing even in between-subjects contexts. A parallel-competitive account thus predicts a negative association between cognitive capacity and the framing effect; it is largely incompatible with an absence of association. These predictions are supported by evidence that asking for a rationale, thus presumably engaging Type 2 thinking, decreases the effect even in between-subjects contexts (McElroy & Seta, 2003; Miller & Fagley, 1991; Takemura, 1994, Experiment 1), while presenting the task under time constraints, thereby preventing Type 2 thinking, increases it (Guo et al., 2017; Takemura, 1994, Experiment 2; but see Igou & Bless, 2007, for findings in the opposite direction).

A more recent subtype of the dual-process framework is the hybrid view (De Neys & Pennycook, 2019), which posits that adherence to rational norms is often due to Type 1 'logical' intuitions rather than deliberate Type 2 thinking. Yet, these logical intuitions result from the automatization of procedures that were originally deliberate (De Neys & Pennycook, 2019), ensuring better intuitions for those of higher cognitive capacity (Thompson et al., 2018). Therefore, the best prediction of the hybrid view would be a negative association between WMC and the framing effect. Unlike rules of logic, however, it is not clear what specific procedures have to be automatized to prevent the between-subjects framing effect. This might be one reason why framing problems are notably absent from empirical tests of the hybrid account, a reason that also makes the account compatible with an absence of association.

1.2.2. Fuzzy-trace theory

A different account for the framing effect is proposed by fuzzy-trace theory (Corbin et al., 2015; Reyna, 2012; Reyna & Brainerd, 1991). Fuzzy-trace theory proposes that people represent information in both verbatim detailed form and gist form, keeping only the essential meaning. In the case of the between-subjects framing effect, verbatim information is typically useless because it leads to the same numbers (e.g., $200 = 1/3 \times 600$). Therefore, people have to rely on gist information. Although different levels of gist representations exist, fuzzy-trace theory predicts that people would rely on the simplest representation sufficient to elicit a preference (Corbin et al., 2015; Reyna, 2012). This lowest-level gist representation is provided by translating the numerical information into the fuzzy categories of 'some' and 'none'. In particular, people would prefer 'saving some' (Program A) to 'saving some vs. saving none' (Program B) in the gain frame. However, they would prefer 'some dying vs. none dying' to 'some dying' in the loss frame.

This natural tendency of using gist representation can be overridden in within-subject contexts, allowing individuals to recognize the equivalence between the different versions of the same problem. The more able individuals are to inhibit the gist-based response, the more likely they will be to exhibit consistent responses (Corbin et al., 2015). Thus, similar to the dual-process accounts and consistent with empirical findings, fuzzy-trace theory predicts that individuals of higher cognitive capacity would resist within-subjects framing better. However, in a between-subjects context, the only helpful interpretation is provided by the gist. Therefore, fuzzy-trace theory predicts no relationship between cognitive capacity and between-subjects framing (Corbin et al., 2015). Fuzzy-trace theory also interprets the between-subjects framing effect as 'an indicator of advanced processing' (Corbin et al., 2015, p. 86), and is thus compatible with Corbin et al.'s (2010) findings for a higher framing effect among the more cognitively able (Corbin et al., 2015; see also Reyna et al., 2014). Indeed, recent independent research has provided evidence that higher WMC is positively associated with both gist and verbatim encoding (Nieznański & Obidziński, 2019).

Table 1. *Competing predictions about the association between cognitive capacity and the between-subjects framing effect.*

		Compatible with a/n . . . association?		
Theory	Reference	Positive	Negative	Absent
Dual-process, default-interventionist	Kahneman (2003); Stanovich and West (2008)	No	No	**Yes**
Dual-process, parallel-competitive	Sloman (1996)	No	**Yes**	No
Dual-process, hybrid	De Neys and Pennycook (2019)	No	**Yes**	Yes
Fuzzy-trace theory	Corbin et al. (2015)	Yes	No	**Yes**
Pragmatic inference	Mandel and Kapler (2018)	No	No	**Yes**
Intensified context shift	Delaney and Sahakyan (2007)	**Yes**	No	No

Note. Best prediction in bold.

1.2.3. Pragmatic accounts

While dual-process accounts and fuzzy-trace theories diverge in many respects, they agree that the framing effect violates normative principles (Chick et al., 2016; Kahneman, 2003). There are alternative approaches that, however, question this interpretation. For instance, Mandel (2014) argued that people presented with the framing tasks interpret some quantifiers as lower bounds. That is, when reading the gain-framed Program A of the disease problem, people infer that *at least* 200 people will be saved. Under this interpretation, Program A is superior to Program B. Similarly, having at least 400 people dying is inferior to Program B. Since this account treats the framing effect as resulting from pragmatic linguistic inferences, it posits that cognitive sophistication is not related to the size of the effect (Mandel & Kapler, 2018).

1.2.4. The intensified context shift hypothesis

Unexpected theoretical support for Corbin et al.'s (2010) findings comes from research on directed forgetting (Delaney & Sahakyan, 2007) which found that high-WMC participants recalled fewer items from a word list than low-WMC participants after an instruction to change the context (i.e., to imagine walking through one's childhood home). Delaney and Sahakyan accounted for the findings with the intensified context shift hypothesis, whereby higher-WMC individuals are better able to access the context in which information was encoded. This advantage, however, comes at the cost of being more dependent on the context during retrieval. Over-reliance on irrelevant context information has also been proposed to be a major mechanism underlying the framing effect (Stanovich et al., 2016). To the extent that encoding and retrieval are also involved in solving a framing problem, high-WMC individuals might be more affected by the contextual information within each frame and thus show a larger framing effect, a prediction also noted by Corbin et al. (2010).

Table 1 summarizes the predictions of the competing accounts. Two accounts are compatible with more than one outcome, but no outcome is compatible with all accounts. Providing a well-powered response to the association question can inform alternative accounts and suggest revising those that are not supported by evidence.

1.3. The present study

To closely replicate Corbin et al.'s (2010) experiment, we used the disease problem (Tversky & Kahneman, 1981) and a version of Ospan (Oswald et al., 2015; Unsworth et al., 2005), which allowed us to directly replicate the original statistically significant frame by Ospan score interaction effect on the response to the disease problem. Although Corbin et al. used the absolute Ospan score, researchers

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

involved in devising the automated complex span tasks recommended the partial score because it has higher internal consistency, correlates better with measures of fluid intelligence, and makes more sense in terms of test theory (Redick et al., 2012).[1] Therefore, we predicted a significant frame by WMC interaction both when using the absolute Ospan score (Hypothesis 1a) and the partial Ospan score (Hypothesis 1b).

To further test the validity and generalizability of the findings, we added the following extensions to the original method. First, to test whether the effect would generalize beyond the disease problem, we included five more risky-choice framing problems (Berthet, 2021; Fagley et al., 2010; Simon et al., 2004). Similarly, we tested if the effect generalizes beyond Ospan. Ospan contains variance from both WMC and the task itself, meaning it also measures factors unrelated to WMC, such as the speed at solving math problems (Foster et al., 2015). In addition, Ospan has been generally found to have smaller loadings on a latent WMC factor compared with other complex span tasks (Draheim et al., 2018). Accordingly, we administered shortened versions of the automatic reading span (Rspan), and automatic symmetry span (SymSpan) and combined their scores into a single WMC score which is more valid than using Ospan alone (Oswald et al., 2015). For our extended replication, we tested if ratings on the framing problems would be significantly predicted by a frame by WMC interaction when a composite WMC task is used (Hypothesis 2).

The Stage 1 registered report, including introduction, method, and planned analyses, was preregistered at the Open Science Framework at November 7, 2002, https://osf.io/grp5m, and updated on April 28, 2023, https://osf.io/hw2sm.[2]

2. Method

2.1. Sampling plan

For the close replication, we performed power simulations by adapting guidelines originally provided for mixed-effects models (DeBruine & Barr, 2021) to the simpler general linear model used in the original study (Corbin et al., 2010). One set of simulations, using parameters extracted from the original data, showed that the original study had a 59% statistical power to detect a significant frame by Ospan interaction.[3] This estimate converged with the estimate of 56% power from analytical power analysis using the MBESS R package (Kelley, 2021).

In another set of simulations, we determined our smallest effect of interest (SESOI) by estimating the effect size that the original study had a 33% power to detect (Simonsohn, 2015). To incorporate predictions from various theoretical accounts, we also considered negative interactions. Results showed that the original study had 33% power to detect betas of the interaction term of sizes −0.033 and +0.034. We then performed a third set of simulations to estimate the sample size needed to detect SESOI with 80% power. To this aim, we varied the number of participants from 200 to 400, in steps of 50, and beta estimates for the interaction term from −0.05 to 0.05, in steps of 0.02 (5,000 simulations per combination). A study with 400 participants would have more than 80% power to detect an interaction equal to ±0.03, which is roughly equal to SESOI (Figure 1).

For the extended replication, we considered using several framing problems. Accordingly, in our simulations, we used mixed-effect modeling that allows using all individual responses in an unaggregated form rather than averaging by participant (Brown, 2021). Mixed-effects modeling is a more flexible analytical approach than linear regression and is related to higher statistical power and lower Type I error rate (Baayen et al., 2008; Barr et al., 2013). In our power simulations

[1] The difference between the absolute and the partial score is described in Method/Materials/WMC Measure/Scoring.

[2] After the Stage 1 registered report had been accepted, we added a block including a 7-item version of the Cognitive Reflection Test (Toplak et al., 2014) at the end of the survey to collect data for an unrelated study (Rachev et al., 2023). We have received the editor's agreement to do so. The updated registration describes that change. Nothing else has been changed relative to the original registration.

[3] The average estimate of the interaction in the simulations was 0.041.

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

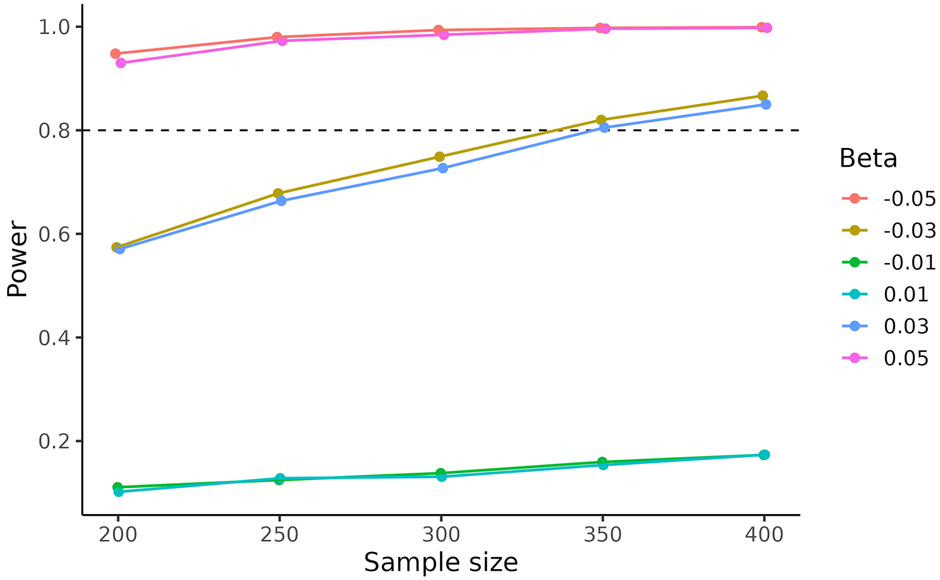

Figure 1. *Close replication: Power to detect the interaction effect at varying sample sizes and values of the interaction term (beta) (5,000 simulations per combination).*

(DeBruine & Barr, 2021), we varied the number of participants from 200 to 400, in steps of 50, the number of framing tasks from three to six, and the beta of the interaction term, in steps of 0.02 (1,000 simulations per combination). The simulations showed that a sample size of 400 participants would ensure 90% power or more to detect SESOI (Figure 2). This number was consistent with the estimates for the close replication. We also performed a Bayes Factor Design Analysis for fixed-*N* designs (Stefan et al., 2019) using the Shiny app the authors provided for the purpose (http://shinyapps.org/apps/BFDA/). Results showed that 200 participants per group would provide at least 85% probability to detect a Bayes factor (BF) of at least 3 (in case of true effect) or 1/3 (in case of null effect) using either informed or default priors and an expected standardized effect size of 0.35.[4] We thus decided to set 400 as our target sample size and six as the number of framing problems. All files to reproduce simulations are available at https://osf.io/ektfd/.

2.2. Participants

Participants were 562 North Americans recruited via Prolific. We used two Prolific workspaces, one funded by Sofia University and the other funded by Florida Gulf Coast University. Participants were compensated by Euro 7.50 at the former workspace and by USD 9.60 at the latter. Five hundred and fifty participants provided full data. One hundred and twenty-five participants were excluded who

- indicated a nationality other than Canada/USA ($n = 0$);
- indicated native language other than English ($n = 0$);

[4]To produce a standardized effect size (Cohen's d), we first calculated a t-value, dividing SESOI, 0.034, by the standard error of the interaction term in the linear model run on Corbin et al.'s data. We then transformed the t-value into d-value using the t_to_d function from the *effectsize* R package (Ben-Shachar et al., 2022). We were unable to enter the exact value 0.38 in the Shiny app, so we opted for the closest smaller value, 0.35.

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

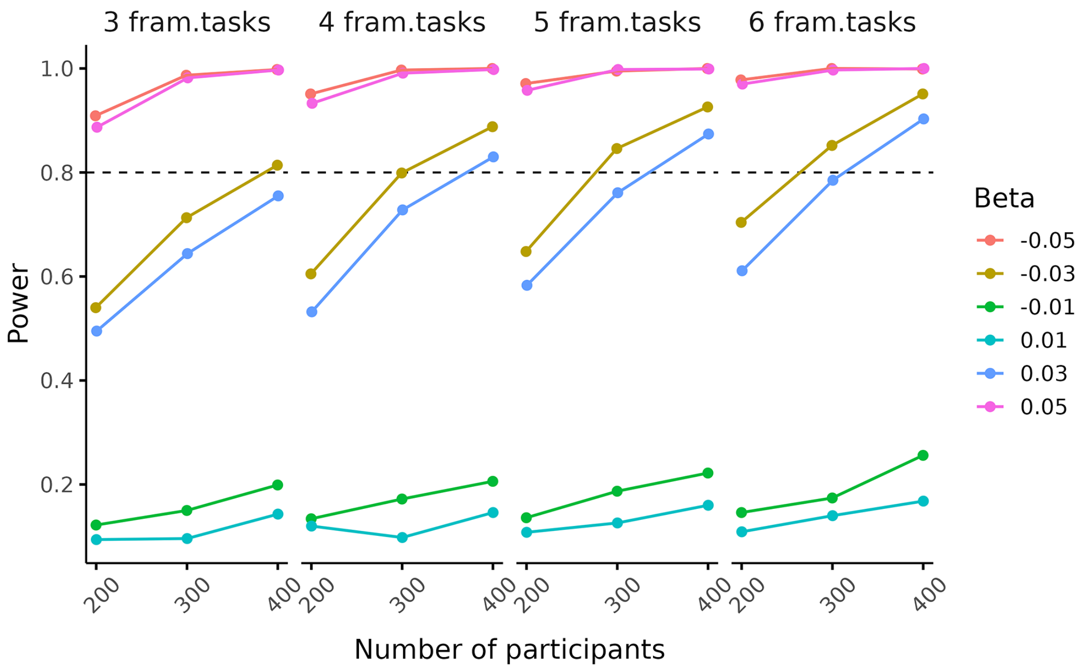

Figure 2. *Extended replication: Power to detect the interaction effect at varying numbers of participants, numbers of framing items, and beta values of the interaction term (1,000 simulations per combination).*

- completed the experiment multiple times, as indicated by identical IP addresses and demographic information (n = 1);
- responded 'yes' to the question asking about the familiarity with the framing tasks (n = 111);
- were less than 95% likely to be above the guessing level on the processing component of the three WMC tasks overall, that is, with overall accuracy below 59.5% (Richmond et al., 2021, and R code provided by their anonymous reviewer) (n = 13).

The final sample consisted of 425 participants (207 female, 214 male, 4 other or prefer not to disclose), mean age 42.56 years (SD = 13.75, Min = 20, Max = 78, Med = 40). Participants were randomly assigned to a gain (n = 222) or loss (n = 203) condition.

2.3. Materials

2.3.1. WMC measure

The shortened versions of Ospan, Rspan, and Symspan (Oswald et al., 2015) were used, as described below. Each of these complex span tasks consists of both a storage component (i.e., items to be recalled) and a processing component (i.e., tasks involving math operations, sentence comprehension, or symmetry judgments). As practice trials for each task, participants went through the storage component alone, the processing component alone, and then the processing component followed by the storage component. For the purpose of the study, the tasks were scripted in PsyToolkit (Stoet, 2010, 2017).

Operation span. Participants were given a set of simple arithmetic operations. For each of them, they had to indicate whether a suggested answer was true or false and were then presented a letter to remember. At the end of each trial, participants were shown a 4×3 letter matrix and asked to click on the letters in the same order as the one they were presented in. Set sizes ranged from 4 to 6, with each set size being administered three times (45 processing-storage trials in total).

Reading span. Participants were asked whether various sentences (of length 10–16 words in English) made sense, with each of them followed by a letter to be remembered and consequently recalled at the end of the set. Set sizes ranged from 4 to 6, with each set size being administered three times (45 trials in total).

Symmetry span. Participants were presented with an 8 ×8 matrix of white and black squares and had to indicate whether the pattern was symmetrical along the vertical axis. After each matrix, the participants were shown a single red square positioned in a 4 × 4 matrix to be remembered and consequently recalled at the end of the set. Set sizes ranged from 3 to 5, with each set size being administered three times (36 trials in total).

Scoring. The complex span tasks are typically scored in two ways. The absolute score is the sum of correctly recalled items in perfectly recalled sets; the partial score is the sum of all correctly recalled items (Unsworth et al., 2005). In line with the original study (Corbin et al., 2010), we used the Ospan absolute score for the close replication (Hypothesis 1a). Yet, as partial scores typically have superior psychometric characteristics than absolute scores (Redick et al., 2012), we also used partial scores as a robustness check in the close replication (Hypothesis 1b) and as the only score in the extended replication (Hypothesis 2).

2.3.2. Framing tasks

Six framing problems were selected after Fagley et al. (2010) and Berthet (2021), namely the disease (adapted from Tversky & Kahneman, 1981), civil defense (adapted from Fagley & Miller, 1997; see also Fischhoff, 1983), cancer treatment (adapted from Fagley & Miller, 1987), traffic accident (adapted from Wang, 1996), African village, and derailed train (adapted from Svenson & Benson III, 1993). These problems have shown good internal consistency in previous research (Berthet, 2021; Fagley et al., 2010). Each problem presented a choice between a sure and a risky option of equal expected values. An example is the disease problem:

> Imagine that the USA/Canada is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:
>
> *(Gain frame)*
>
> If program A is adopted, 200 people will be saved.
>
> If program B is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved.
>
> *(Loss frame)*
>
> If program A is adopted, 400 people will die.
>
> If program B is adopted, there is a 1/3 probability that nobody will die and a 2/3 probability that 600 people will die.
>
> Which of the two programs would you favor?

All the problems were related to human lives to keep constant the potentially important factor of task content in risky-choice framing problems (Fagley & Miller, 1997). Following Fagley et al.'s (2010) recommendations, the number of lives at risk was constrained to be on roughly the same scale (varying from 100 to 600). The chances for no one saved in the probabilistic option were 2/3 in half the problems and 3/4 in the other half. In line with Corbin et al. (2010), participants indicated their preference on each task on a 7-point scale ranging from 1 (definitely would favor A) to 7 (definitely would favor B). Option

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

A was always the safe option and Option B was always the risky option. The problems are listed in the Appendix.

2.3.3. Familiarity with the framing tasks

Since familiarity with the framing tasks reduces the framing effect (Rachev et al., 2021), participants were asked: ‘Have you seen any of these choice problems before?’, with response options ‘yes’ and ‘no’.

2.4. Procedure

Data were collected online using PsyToolkit (Stoet, 2010, 2017). Participants provided informed consent and filled in demographic information (nationality, gender, age, educational status, and major). They then completed the complex span tasks and the framing problems, presented in separate blocks. Unlike Corbin et al. (2010) who used a fixed order with Ospan always coming first, we counterbalanced the order of the WMC and framing blocks. This was done to prevent artifacts of task order (e.g., individuals with lower WMC might have been more exhausted by Ospan and consequently paid less attention to the framing problem, resulting in a smaller framing effect among these participants). To ensure close replication, Ospan and the disease problem were always kept first in their blocks, thus preventing carryover effects.

Within the WMC block, Ospan was followed by Rspan and then by Symspan. Within the Framing block, the five tasks following the disease problem were presented in a random order. We also exploratively checked if the problems were successful in eliciting a framing effect within subjects, as opposed to evoking a risk-seeking or risk-averse attitude regardless of the frame. To this aim, after rating the six problems in their assigned frame, participants were asked to rate two of the problems in the alternative frame.[5] Participants were explicitly told that the aim was to check if it mattered to them whether the problems are stated as gains or losses. Immediately after the last framing problem came the familiarity question. The median completion time was 49 min.

Table 2 summarizes the differences between Corbin et al.’s (2010) method and our method.

2.5. Piloting

To test if the script worked as intended, we ran a pilot study on ($N = 30$) participants. It also served as a pre-test for the framing and WMC measures. No issue has arisen during the piloting. The pilot dataset and summary statistics are available at OSF, https://osf.io/ektfd/.

3. Results

3.1. Summary statistics

Means, standard deviations, and intercorrelations are displayed for the framing tasks (Table 3) and WMC tasks (Table 4).

3.2. Measures’ psychometric properties

Framing problems: Internal consistency. The internal consistency of the framing problems was Cronbach’s alpha = 0.87, which is higher than the values reported in our reference studies (Berthet, 2021, Study 2; Fagley et al., 2010; Simon et al., 2004).

[5]We were unable to randomly pick two problems in PsyToolkit. Rather, we randomly assigned participants to one out of three possible pairs exhausting the total of six framing problems.

Table 2. *Differences between Corbin et al.'s method and our method.*

Aspect	Corbin et al. (2010)	Our study	Rationale
Framing measure	Disease problem (Tversky & Kahneman, 1981)	Six framing problems (see 'Materials' section and Appendix for details)	A more reliable measure of risky-choice framing
WMC measure	Automated operation span (Unsworth et al., 2005)	Shortened[a] versions of the automated operation, reading, and symmetry span tasks (Oswald et al., 2015)	A more reliable measure of WMC
WMC scoring	Absolute score	Absolute score only for the closest replication; partial score for extended analyses	More reliable, valid, and informative scoring
Block order	Fixed order	WMC and framing blocks counterbalanced	Presentation order is potentially an important confound
Environment	Lab	Online	COVID-19-related restrictions; easy access to a larger sample

[a]The full-length Ospan task (Unsworth et al., 2005), as used by Corbin et al. (2010), administers three trials with set sizes ranging from 3 to 7, while the shortened version includes two trials with set sizes ranging from 4 to 6.

Table 3. *Mean ratings, standard deviations, and correlations for the framing tasks.*

	Gain ($N = 222$)		Loss ($N = 203$)		Correlations				
	M	SD	M	SD	1.	2.	3.	4.	5.
1. Disease	3.14	1.99	4.51	2.06	–				
2. Civil defense	3.63	1.84	4.46	1.97	0.51	–			
3. Cancer treatment	3.33	1.91	4.58	1.91	0.46	0.48	–		
4. Traffic accident	3.45	1.93	4.24	1.93	0.47	0.56	0.49	–	
5. African village	3.32	1.87	3.98	1.97	0.49	0.51	0.52	0.55	–
6. Derailed train	3.16	1.82	4.33	1.99	0.57	0.56	0.48	0.57	0.54

Note: Ratings ranged from 1 (definitely would favor safe option) to 7 (definitely would favor risky option).

Table 4. *Means, standard deviations, and correlations for the WMC tasks.*

	Absolute score		Partial score		Correlations		
	M	SD	M	SD	1.	2.	3.
1. Operation span	30.74	12.67	36.7	9.89	*0.77/0.83*	0.66	0.42
2. Reading span	31.6	13.44	37.6	9.9	0.66	*0.82/.87*	0.38
3. Symmetry span	12.66	10.23	20.46	9.44	0.44	0.47	*0.78/0.84*

Note. Alpha coefficients in italics on the main diagonal (absolute/partial score), calculated over the recall scores in the three separate blocks (Redick et al., 2012, p. 169). Correlations between the absolute (above the main diagonal) and partial scores (below the main diagonal).

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

Table 5. *Confirmatory analyses: Overview.*

Named analysis	Variation	R formula notation (simplified)	AIC
Close replication	Original	Rating_DP ~ Frame * Ospan_Absolute	
	Robustness check	Rating_DP ~ Frame * Ospan_Partial	
Extended replication	Item-slope	FramRating ~ Frame * WMC_Partial + (Frame | framing_item) + (1 | Participant)	9,736.5
	Independent slope	**FramRating ~ Frame * WMC_Partial + (Frame || framing_item) + 1 | Participant)**	**9,734.7**
	Random intercepts	FramRating ~ Frame * WMC_Partial + (1 | framing_item) + (1 | Participant)	9,742.4

Note. Model in bold is the best model according to AIC (Seedorff et al., 2019).

Working memory capacity: Internal consistency. The internal consistency of the Ospan partial score, calculated over the proportion of correctly recalled letters in the nine trials,[6] was Cronbach's alpha = 0.86, which is higher than the value of 0.71 reported by Oswald et al. (2015) and deemed very good (Greiff & Allen, 2018). The internal consistency of the composite WMC measure, calculated over the total partial scores across the three complex span tasks, was Cronbach's alpha = 0.77, which is roughly the same as the value of 0.76 reported by Oswald et al. (2015) and deemed 70: good (Greiff & Allen, 2018). We used the latter measure to test Hypothesis 2.

Working memory capacity: Confirmatory factor analysis. To investigate the construct validity of the WMC measure, we conducted confirmatory factor analysis (CFA) of the number correct in each trial, assuming one factor. As multivariate normality was violated, we used robust estimators of model fit. The chi-square test indicated no exact fit, $\chi^2(321) = 378.97$, $p = 0.014$. However, approximate fit indices (RMSEA = 0.023, SRMR = 0.036) and incremental fit indices (CFI = 0.984, TLI = 0.983, IFI = 0.978) were excellent, indicating that the hypothesized model fit the data well (Greiff & Allen, 2018, Table 1). Given the good internal consistency of the WMC measure and the excellent fit of the CFA model, we used the full WMC measure as planned.

3.3. Confirmatory analyses

3.3.1. Analytic approach

For the close replication (Hypotheses 1a and 1b), we used linear regression predicting the ratings on the disease problem by the frame (gain/loss), the Ospan score, and the frame by Ospan interaction (Table 5). A statistically significant interaction showing a larger framing effect on higher levels of WMC was to be treated as a successful replication.

For the extended replication (Hypothesis 2), we used mixed-effects modeling (MELM) to predict the ratings on the framing tasks by the frame (gain/loss), the WMC score, and the frame by WMC interaction. To infer the random-effects structure of the data, we compared the three plausible models, namely a model with slopes varying by framing items, an independent slopes model excluding the correlation between by-item varying slopes and intercepts, and a model that only included random intercepts for items and participants (Table 5). We selected the best model based on AIC and AICc (Seedorff et al., 2019). Then we compared the best model to a model excluding the interaction term.

[6]Redick et al. (2012, p. 169) describe two methods of calculating the internal consistency of complex span task scores. We used the first method in Table 3 and the second method in this paragraph. Note that the second method cannot be used to calculate the internal consistency of the absolute scores as it treats the proportion correct in each trial as 100% or 0%.

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

A significant likelihood ratio test (Winter, 2019) favoring the model that included the interaction term would be treated as a successful replication.

To further investigate the relative evidence in favor of the effect versus the null hypothesis, we calculated BF using the *brms* R package (Bürkner et al., 2020). Unlike frequentist analysis, BF allows inferring not only absence of evidence but also evidence of absence. In other words, it can inform if a null result supports the null hypothesis or is due to the insensitivity of the data (Dienes, 2014). We modeled the alternative hypothesis to have a normal distribution with a mean equal to positive SESOI = 0.034 (or negative SESOI if the effect turned out to be negative) and SD = 1/2 × SESOI (Dienes, 2014). We used default priors for the remaining effects.

In all models, we centered the WMC scores and used a deviation coding for the frame (gain = −0.5, loss = +0.5). These transformations do not qualitatively change the results but reduce non-convergence issues and make it easier to interpret interactions (Brown, 2021; DeBruine & Barr, 2021; Winter, 2019).

All analyses were conducted in R (R Core Team, 2021) and the following packages: afex (Singmann et al., 2023), beepr (Bååth & Dobbyn, 2018), brms (Bürkner et al., 2020), broom (Robinson et al., 2023), broom.mixed (Bolker & Robinson, 2022), car (Fox et al., 2023), cowplot (Wilke, 2020), faux (DeBruine et al., 2023), foreign (R Core Team et al., 2023), ggeffects (Lüdecke et al., 2023), lavaan (Rosseel et al., 2020), lme4 (Bates et al., 2020), MASS (Ripley et al., 2023), MBESS (Kelley, 2021), and tidyverse (Wickham & RStudio, 2019). The R code is available at https://osf.io/ektfd/.

3.3.2. Close replication

For the close replication, we used the disease problem and the Ospan absolute score (Hypothesis 1a). As a robustness check, we also analyzed the data using the Ospan partial score (Hypothesis 1b). Table 5 displays all the confirmatory analyses using R formula notation. Table 6 displays the estimates from these models.

Manipulation check. To check if the framing of the disease problem was effective, we performed linear regression with ratings on the disease problem as the outcome variable and frame as a single predictor. As expected, the frame significantly predicted ratings, beta = 1.37, $t = 6.98$, $p < 0.001$, confirming that the framing manipulation was successful.

Assumption checks. We also checked the assumptions underlying multiple linear regression, namely normality, linearity, homoscedasticity, and collinearity (Winter, 2019). The normality assumption was violated in the models used to test both Hypotheses 1a and 1b.[7] Bootstrapping the regression model (Fox & Weisberg, 2021) over 10,000 samples confirmed the results from the regression model (Table 6). To test Hypotheses 1a and 1b, we used the parameters estimated from the ordinary least squares regression.

Hypothesis 1a: Frequentist test. In the model using the Ospan absolute score, the critical interaction term, $b = 0.002$ (Table 6), was very close to 0, and the 95% confidence interval, [−0.03, 0.03], crossed 0, meaning no substantial change in the framing effect as WMC increases (Figure 3A). This effect is descriptively inconsistent with Corbin et al.'s (2010) and is not statistically significant, $p = 0.89$. Hypothesis 1a was therefore not supported.

Hypothesis 1a: Bayes factor. The BF, BF_{10} = 0,27, shows that the data are 3.75 times more consistent with a model omitting the interaction term than with a model including it, meaning moderate evidence in favor of the null hypothesis of no interaction over Hypothesis 1a (Stefan et al., 2019, Table 1).

Hypothesis 1b: Frequentist test. In the model using the Ospan partial score, the critical interaction term, $b = 0.007$ (Table 6), was very close to 0, and the 95% confidence interval, [−0.04, 0.04], crossed 0, meaning no substantial change in the framing effect as WMC increases (Figure 3B). This effect is descriptively inconsistent with Corbin et al.'s (2010) and is not statistically significant, $p = 0.95$. Hypothesis 1b was therefore not supported.

[7]For details, see https://osf.io/ektfd/, file '07B_ResultsH1.html'.

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

Table 6. *Confirmatory analyses: Model estimates.*

		Close replication				Extended replication
		Ospan absolute		Ospan partial		
		OLS	Bootstrap	OLS	Bootstrap	
Fixed effects						
Intercept	Beta	3.82***	3.82	3.82***	3.82	3.84
	95% CI	[3.63, 4.01]	[3.60, 4.02]	[3.63, 4.02]	[3.62, 4.04]	[3.67, 4.02]
Frame	Beta	1.36***	1.44	1.36 ***	1.44	1.01
	95% CI	[0.98,1.76]	[0.91, 1.86]	[0.98, 1.75]	[0.99, 1.97]	[0.66, 1.36]
WMC	Beta	0.009	0.009	0.007	0.006	0.00005
	95% CI	[−0.01, 0.02]	[−0.006, 0.025]	[−0.01, 0.03]	−0.02, 0.03]	[−0.006, 0.006]
Frame × WMC	Beta 95% CI	0.002 [−0.03, 0.03]	0.004 [−0.03, 0.4]	−0.001 [−0.04, 0.04]	0.0004 [−0.04, 0.04]	−0.004 [−0.016, 0.008]
Random effects						
Intercept for participants	SD 95% CI					1.35 [1.24, 1.46]
Intercept for framing item[a]	SD 95% CI					0.19 [0.05, 0.27]
Slope for framing item	SD 95% CI					0.24 [0.10, 0.55]

Abbreviations: OLS, ordinary least squares regression; fixed effects: beta, regression coefficient (WMC has been centered, Frame has been deviation coded, i.e. gain = −.5, loss = +.5); random effects: SD, standard deviation. *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

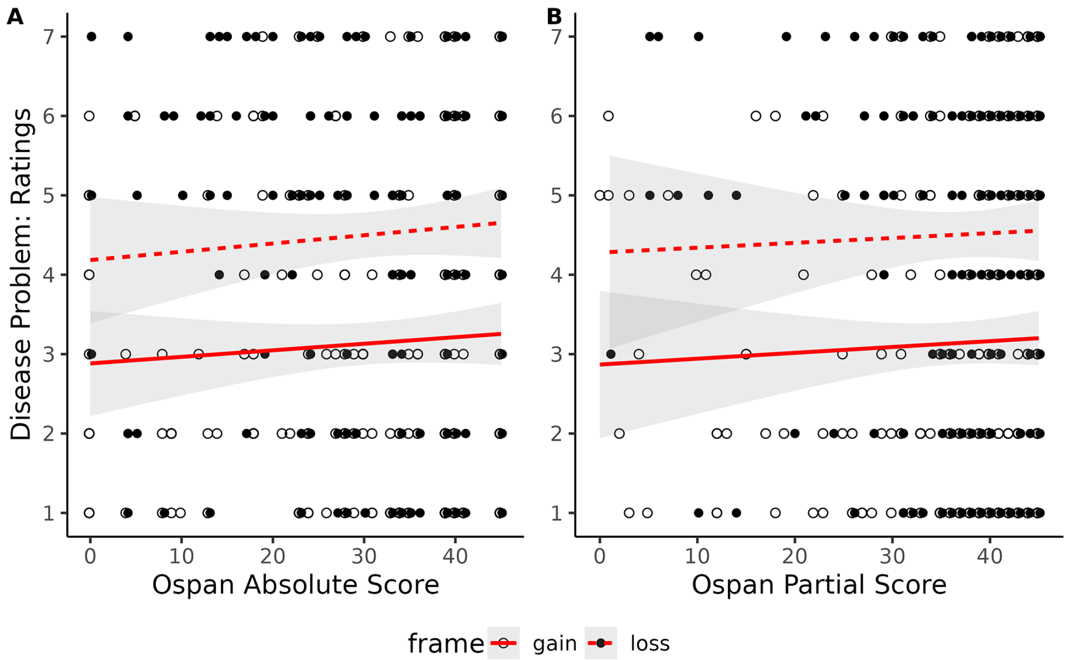

Figure 3. *Close replication: Observed (points) and predicted (lines) framing ratings as a function of WMC (A: Absolute Score, B: Partial Score).*

Hypothesis 1b: Bayes factor. The BF, $BF_{10} = 0.32$, shows that the data are 3.10 times more consistent with a model omitting the interaction term than with a model including it, meaning moderate evidence in favor of the null hypothesis of no interaction over Hypothesis 1b (Stefan et al., 2019, Table 1).

3.3.3. Extended replication

Model selection. No models showed singular fit or convergence issues. The best model in terms of AIC was the independent slope model (Table 5).

Manipulation check. To check if the framing manipulation was effective overall, we compared two mixed-effects linear models with the same outcome variable and random-effects structure as the best model. A model including the frame as a single fixed predictor fit the data better than the model without the frame (i.e., only including the fixed intercept), $\chi^2(1) = 15.21$, $p < 0.001$, confirming that the framing manipulation was successful.

Assumption checks. [IF all LM assumptions met:] All assumptions underlying linear models were met satisfactorily in the best model.[8]

Hypothesis 2: Test. In the best model, the critical interaction term, $b = -0.004$ (Table 6), was very close to 0, and the 95% confidence interval, [−0.016, 0.008], crossed 0, meaning no substantial change in the framing effect as WMC increases (Figure 4). This effect is descriptively inconsistent with Corbin et al.'s (2010). Moreover, a likelihood ratio test indicated that the model including the frame × WMC interaction did not fit the data better than a model without the interaction, $\chi^2(1) = 0.48$, $p = 0.49$. Hypothesis 2, that higher WMC predicts a larger framing effect more generally, was, therefore, not supported.

[8]For details, see https://osf.io/ektfd/, file '07C_ResultsH2.html'.

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

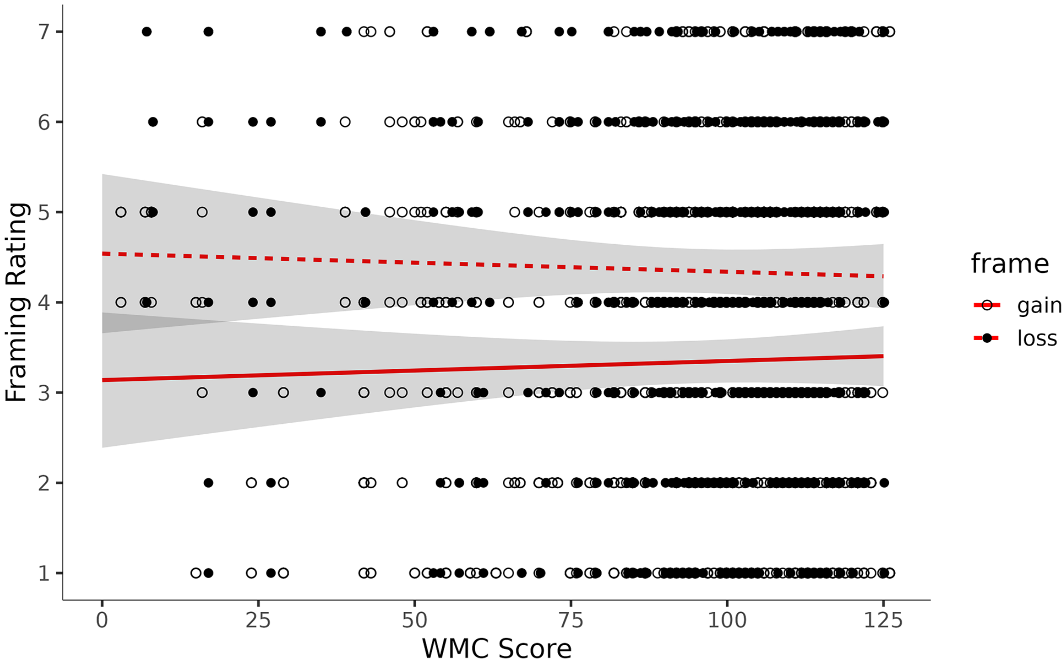

Figure 4. *Extended replication: Observed (points) and predicted (lines) framing ratings as a function of WMC.*

Hypothesis 2: Bayes factor. The BF, $BF_{10} = 0.047$, shows that the data are 21.00 times more consistent with a model omitting the interaction term than with a model including it, meaning strong evidence in favor of the null hypothesis of no interaction over Hypothesis 2 (Stefan et al., 2019, Table 1).

3.4. Exploratory analyses

To check if the problems were successful in eliciting a framing effect within subjects, we subtracted the score in the gain frame from the score in the loss frame for problems presented in both frames, and ran several mixed-effects linear models that included the difference score as the outcome variable and the frame as a single fixed predictor. The best-fitting model was the one including only random intercepts for participants and framing problems. It fit the data significantly better than a model that only included the fixed intercept, $\chi^2(1) = 50.00$, $p < 0.001$, confirming that the problems successfully elicited a framing effect within subjects.

We also checked whether the difference scores are related to WMC by running mixed-effects linear models that included the difference score as the outcome variable and the frame, the WMC score, and their interaction as fixed predictors. The best model, which contained only random intercepts for participants and framing problems, did not fit the data significantly better than the model omitting the interaction term, $\chi^2(1) = 0.75$, $p = 0.39$. In short, there was no evidence within this study that WMC is related to the within-subjects framing effect.

4. Discussion

Corbin et al. (2010) found a larger risky-choice framing effect among people with higher compared with lower WMC. This finding is unusual because most past research failed to find an association between cognitive capacity and the between-subjects framing effect (Stanovich & West, 2008). Whenever any association has been found between cognitive capacity and thinking problems more generally, it was that higher capacity was positively rather than negatively associated with normative performance. Given the potentially important theoretical implications of the association between WMC and the risky-choice framing effect, we set out to replicate and extend it in a registered report, using both the original measures and composite measures of framing and WMC.

In contrast to the original findings, no significant frame by Ospan interaction was found in predicting ratings on the disease-framing problem (Tversky & Kahneman, 1981), using either the absolute or the partial Ospan score. BFs showed moderate evidence in favor of the null hypothesis, thus corroborating the frequentist findings. The close replication was thus unsuccessful.

We were also interested in going beyond the particular tasks used by Corbin et al. (2010), so we extended the materials using six framing tasks related to human lives and three WMC tasks with different processing components (numeric, verbal, and spatial). Consistent with the close replication, the combined WMC score did not significantly interact with the frame to predict the ratings on the framing problems. The BF also showed strong evidence in favor of the null. The extended replication was thus unsuccessful.

Taken together, the present findings converge in casting doubt on the hypothesis of greater susceptibility to framing effects among individuals of higher WMC. All six preregistered tests, including three frequentist and three Bayesian ones, point in the same direction of no association. Such coherent and unequivocal empirical evidence entails a proportionally strong conclusion that the between-subjects risky-choice framing effect is not associated with WMC.

In light of the present unequivocal evidence, we see two main reasons for the positive association originally found by Corbin et al. (2010). Most probably, this was a false positive result, which is common to underpowered studies (e.g., Chambers, 2017). Alternatively, the positive findings might reflect a genuine effect which is, however, due to some third factor that was not accounted for in the original study. Recall that Corbin (2013) only replicated the interaction effect among participants of low numeracy, which might also be the case for Corbin et al.'s (2010) sample. Such an intriguing three-way interaction between WMC, numeracy, and the framing effect might be theoretically defendable (e.g., from the point of view of the fuzzy-trace theory), but testing it would require a sample roughly four times as big as in the present study (Simonsohn, 2014), that is, about 1,600 participants. Until a good rationale and resources are found to test this possibility, we will content ourselves with concluding that WMC is not related, overall, to the between-subjects risky-choice framing effect.

4.1. Our findings in the context of extant empirical evidence

Our findings are consistent with previous research that failed to find association between cognitive capacity and performance on a host of decision-making tasks administered between-subjects, including framing problems (Stanovich & West, 2008). Our findings diverge from evidence for an advantage of higher WMC in decision-making tasks (Burgoyne et al., 2023; Cokely & Kelley, 2009; Del Missier et al., 2013; Dougherty & Hunter, 2003; Starcke et al., 2011). There are many differences between the problems used in those studies and the framing problems we used that could account for this divergence. Most notably, superior performance associated with higher WMC appears in tasks that demonstrate an individual's reasoning capacity to adhere to rational norms. In framing tasks presented between subjects, on the other hand, there is neither a clear procedure nor domain-specific knowledge that the individual problem solver could follow in order to adhere to the rational norm because the latter is only tested at the supra-individual level, as coherence of people's beliefs and preferences. This distinction between reasoning rationality and coherence rationality is important in many respects and the former does not necessarily imply the latter (Kahneman, 2000). Better access to highly automatized procedures, which is supposedly at the core of the WMC advantage, is of no use in the absence of clear rules to be followed. Therefore, even high-WMC individuals are expected to fail the stricter coherence-rationality test even though they are capable of passing less strict reasoning-rationality tests.

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

Our exploratory analyses also failed to find an association between WMC and the within-subjects framing effect we found, while a negative association has sometimes emerged in previous research (Bruine de Bruin et al., 2007), although not always (Del Missier et al., 2012; Stanovich & West, 1998; Toplak et al., 2014). Our primary goal behind presenting some tasks in the alternative frame was to check if the problems successfully elicited the framing effect, which they did. We were quite explicit about this aim in our instruction to participants. By contrast, previous research has typically introduced the alternative frame without an extra explanation, using all the problems from the original frame, and with a certain time interval between the two frames filled with other tasks. Given these differences, we do not consider our results to be a genuine test of a within-subjects framing effect. Consequently, we could not rigorously test for an association between WMC and the within-subjects framing effect, although we think that such a test would further advance the evaluation of the competing accounts of the framing effect.

4.2. Theoretical implications

The outcome of this study turned out to be both the most expected one empirically and also the least informative one theoretically. Four out of the six accounts we considered were compatible with the present outcome of no association between WMC and the between-subjects framing effect, while the two alternative outcomes (positive and negative association) had only two supporting accounts each (Table 1). Still, we can conclude that the findings are incompatible with two accounts, the parallel-competitive dual-process account (Sloman, 1996) and the intensified shift account (Delaney & Sahakyan, 2007). Further, we may want to assign more credit to accounts that specifically predicted no association as the single compatible outcome than to accounts that were also compatible with alternative outcomes. By this logic, the more specific default-interventionist (Kahneman, 2003; Stanovich & West, 2008) and pragmatic inference (Mandel & Kapler, 2018) accounts predicted the outcome better than the vaguer fuzzy-trace (Corbin et al., 2015) and the hybrid dual-process (De Neys & Pennycook, 2019) accounts.

To be sure, some might not agree with our interpretations of the various theoretical predictions, which are in some cases our best inferences rather than claims explicitly made by the authors of the respective theories. Such disagreement about what theories entail is a common problem in psychology where theories are not axiomatized, but it is the theories that are to blame rather than those who tested them (Evans, 2016). The burden is then on the authors to further elaborate their theories to the point that they make clear predictions that could be logically derived and empirically tested by independent researchers.

4.3. Limitations and avenues for future research

Our study had several methodological limitations. First, we used only risky-choice framing problems related to human lives. This might be the reason why the internal consistency of the framing problems was higher than usual. However, research has shown that the risky-choice framing effect is sometimes moderated by the task contents (Fagley & Miller, 1997). Future research might test whether the present findings generalize to risky-choice problems in the financial domain or to other types of framing effects such as the attribute framing effect (Levin et al., 1998).

Second, multiple definitions and measures of working memory exist (Cowan, 2017; Oberauer et al., 2018), so the association between WMC and the framing effect can be further tested using other measures of WMC. Third, the study was conducted online and thus lacked rigorous procedural control. Yet, our quality checks and descriptive statistics did not point to anything that would noticeably affect participants' performance, so we believe the upside of online data collection, in terms of fast and easy recruitment of participants beyond student samples, outweighs its downside in this case.

Since our primary goal was to replicate and extend Corbin et al.'s (2010) findings, we missed opportunities to pit against each other the two dominant accounts of the present findings, the

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

default-interventionist dual-process account and the pragmatic inferences account. For instance, we considered adding the word ‘exactly’ to the numbers of people affected (e.g., ‘exactly 200 people will be saved’). According to the pragmatic inference account, adding ‘exactly’ would prevent people from interpreting the number as lower bound (‘at least’) and will thus make the framing effect disappear, but according to the default-interventionist account, this single-word addition would not affect the framing effect. However, such an addition would also impede close replication. Hence, we refrained from changing the original wording of the problems. Likewise, we have refrained from testing for a WMC by framing interaction in a within-subjects design, which would be another way to distinguish between the two accounts.[9] We look forward to future research aimed at assessing alternative explanations through refined methods. The evidence we have presented can be combined with forthcoming findings to better distinguish between the various theories of framing and its moderating factors.

Acknowledgments. We thank Maria Ilieva for her contribution in the early stages of the project; Ekaterina Peycheva, Julia Kamburidis, and Evelina Marinova for the translation of materials into Bulgarian; Nancy Fagley for sharing the materials for the Fagley et al.’s (2010) study, and Jason Tsukahara and the Attention & Working Memory Lab at Georgia Institute of Technology for sharing with us the E-Prime versions of the Oswald shortened complex span tasks and for responding to our queries. We also thank Kiril Kostov and the Department of Cognitive Science and Psychology at the New Bulgarian University for their willingness to provide us with an internet environment to run the tasks, had we used E-Prime. Finally, we thank Gijsbert Stoet for kindly responding to all our queries related to scripting the materials in PsyToolkit.

Author contributions. Conceptualization: B.B., J.C., S.D., T.M., N.R.R.; Data curation: N.R.R.; Formal analysis: N.R.R.; Funding acquisition: T.M., N.R.R.; Investigation: B.B., S.D., T.M., N.R.R.; Methodology: B.B., S.D., N.R.R.; Project administration: N.R.R.; Resources: N.R.R.; Software: N.R.R.; Supervision: N.R.R.; Validation: N.R.R.; Writing—original draft: B.B., S.D., N.R.R.; Writing—review and editing: B.B., J.C., S.D., T.M., N.R.R. The authors have agreed to list their names alphabetically (Chambers, 2017).

Data availability statement. Data for this study are available at https://osf.io/ektfd/.

Funding statement. This research was funded by the Sofia University Scientific Research Fund, Contract #21/2016, awarded to N.R.R., and by a grant by Florida Gulf Coast University awarded to T.M.

Competing interest. We have no relevant competing interests to disclose.

References

Bååth, R., & Dobbyn, A. (2018). *beepr: Easily Play Notification Sounds on any Platform* (1.3) [Computer software]. https://cran.r-project.org/web/packages/beepr/index.html

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., Walker, S., & Krivitsky, P. N. (2020). *lme4: Linear mixed-effects models using “Eigen” and S4* (1.1-25) [Computer software]. https://CRAN.R-project.org/package=lme4

Ben-Shachar, M. S., Makowski, D., Lüdecke, D., Indrajeet Patil, Wiernik, B. M., Kelley, K., Stanley, D., Burnett, J., & Karreth, J. (2022). *effectsize: Indices of effect size and standardized parameters* (0.7.0) [Computer software]. https://CRAN.R-project.org/package=effectsize

Berthet, V. (2021). The measurement of individual differences in cognitive biases: A review and improvement. *Frontiers in Psychology*, *12*, 419. https://doi.org/10.3389/fpsyg.2021.630177

Bolker, B., & Robinson, D. (2022). *broom.mixed: Tidying methods for mixed models* (0.2.9.4) [Computer software]. https://CRAN.R-project.org/package=broom.mixed

Brevers, D., Cleeremans, A., Goudriaan, A. E., Bechara, A., Kornreich, C., Verbanck, P., & Noël, X. (2012). Decision making under ambiguity but not under risk is related to problem gambling severity. *Psychiatry Research*, *200*(2), 568–574. https://doi.org/10.1016/j.psychres.2012.03.053

[9]The default-interventionist account would predict a smaller within-subjects framing effect among individuals with higher WMC, because they are better able and willing to be consistent with their initial responses (i.e., they are better in terms of reasoning rationality). The pragmatic account, on the other hand, would predict no association, just as in a between-subjects design, because participants would not consider the problems to be equivalent at the first place.

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920960351. https://doi.org/10.1177/2515245920960351
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. https://doi.org/10.1037/0022-3514.92.5.938
Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S., Hambrick, Z., & Engle, R. W. (2023). Understanding the relationship between rationality and intelligence: A latent-variable approach. *Thinking & Reasoning*, *29*, 1–42.
Bürkner, P.-C., Gabry, J., & Weber, S. (2020). *brms: Bayesian regression models using "Stan"* (2.14.4) [Computer software]. https://CRAN.R-project.org/package=brms
Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice* (1st ed.). Princeton, NJ: Princeton University Press.
Chick, C. F., Reyna, V. F., & Corbin, J. C. (2016). Framing effects are robust to linguistic disambiguation: A critical test of contemporary theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 238–256. https://doi.org/10.1037/xlm0000158
Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, *4*(1), 20–33.
Corbin, J. C. (2013). *Unbounded irrationality: Memory, individual differences, framing effects, and fuzzy-trace theory* [Master's thesis, Cornell University]. https://ecommons.cornell.edu/handle/1813/34259
Corbin, J. C., Liberali, J. M., Reyna, V. F., & Brust-Renck, P. G. (2015). Intuition, interference, inhibition, and individual differences in fuzzy-trace theory. In E. A. Wilhelms & V. F. Reyna (Eds.), *Neuroeconomics, judgment, and decision making* (pp. 77–90). New York: Psychology Press.
Corbin, J. C., McElroy, T., & Black, C. (2010). Memory reflected in our decisions: Higher working memory capacity predicts greater bias in risky choice. *Judgment and Decision Making*, *5*(2), 110–115.
Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6
De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, *28*(5), 503–509. https://doi.org/10.1177/0963721419855658
DeBruine, L. M., & Barr, D. J. (2021). Understanding mixed-effects models through data simulation. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920965119. https://doi.org/10.1177/2515245920965119
DeBruine, L. M., Krystalli, A., & Heiss, A. (2023). *faux: Simulation for factorial designs* (1.2.1) [Computer software]. https://cran.r-project.org/web/packages/faux/index.html
Del Missier, F., Mäntylä, T., & Bruine de Bruin, W. (2012). Decision-making competence, executive functioning, and general cognitive abilities. *Journal of Behavioral Decision Making*, *25*(4), 331–351. https://doi.org/10.1002/bdm.731
Del Missier, F., Mäntylä, T., Hansson, P., Bruine de Bruin, W., Parker, A. M., & Nilsson, L.-G. (2013). The multifold relationship between memory and decision making: An individual-differences study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1344.
Delaney, P. F., & Sahakyan, L. (2007). Unexpected costs of high working memory capacity following directed forgetting and contextual change manipulations. *Memory & Cognition*, *35*(5), 1074–1082. https://doi.org/10.3758/BF03193479
Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781
Dougherty, M. R. P., & Hunter, J. (2003). Probability judgment and subadditivity: The role of working memory capacity and constraining retrieval. *Memory & Cognition*, *31*(6), 968–982. https://doi.org/10.3758/BF03196449
Draheim, C., Harrison, T. L., Embretson, S. E., & Engle, R. W. (2018). What item response theory can tell us about the complex span tasks. *Psychological Assessment*, *30*(1), 116–129. https://doi.org/10.1037/pas0000444
Evans, J. S. B. T. (2016). *How to be a researcher: A strategic guide for academic success* (pp. xi, 162). New York: Routledge/Taylor & Francis Group.
Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685
Fagley, N. S., Coleman, J. G., & Simon, A. F. (2010). Effects of framing, perspective taking, and perspective (affective focus) on choice. *Personality and Individual Differences*, *48*(3), 264–269.
Fagley, N. S., & Miller, P. M. (1987). The effects of decision framing on choice of risky vs certain options. *Organizational Behavior and Human Decision Processes*, *39*(2), 264–277.
Fagley, N. S., & Miller, P. M. (1997). Framing effects and arenas of choice: Your money or your life? *Organizational Behavior and Human Decision Processes*, *71*(3), 355–373. https://doi.org/10.1006/obhd.1997.2725
Fischhoff, B. (1983). Predicting frames. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 103–116. https://doi.org/10.1037/0278-7393.9.1.103
Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–236. https://doi.org/10.3758/s13421-014-0461-7
Fox, J., & Weisberg, S. (2021). *Bootstrapping regression models in R: An appendix to an R companion to applied regression* (3rd ed.) (p. 18). https://socialsciences.mcmaster.ca/jfox/Books/Companion/appendices/Appendix-Bootstrapping.pdf

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Krivitsky, P., Laboissiere, R., Maechler, M., Monette, G., Murdoch, D., Nilsson, H., . . . R-Core. (2023). *car: Companion to applied regression* (3.1-2) [Computer software]. https://cran.r-project.org/web/packages/car/index.html

Greiff, S., & Allen, M. S. (2018). EJPA introduces registered reports as new submission format [Editorial]. *European Journal of Psychological Assessment*, *34*(4), 217–219. https://doi.org/10.1027/1015-5759/a000492

Guo, L., Trueblood, J. S., & Diederich, A. (2017). Thinking fast increases framing effects in risky decision making. *Psychological Science*, *28*(4), 530–543. https://doi.org/10.1177/0956797616689092

Igou, E. R., & Bless, H. (2007). On undesirable consequences of thinking: Framing effects as a function of substantive processing. *Journal of Behavioral Decision Making*, *20*(2), 125–142. https://doi.org/10.1002/bdm.543

Kahneman, D. (2000). A psychological point of view: Violations of rational rules as a diagnostic of mental processes. *Behavioral and Brain Sciences*, *23*(05), 681–683.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kelley, K. (2021). *MBESS: The MBESS R package* (4.8.1) [Computer software]. https://CRAN.R-project.org/package=MBESS

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., . . . Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology*, *45*(3), 142–152. https://doi.org/10.1027/1864-9335/a000178

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*(1), 23–55. https://doi.org/10.1006/obhd.1998.2781

Lauriola, M., & Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences*, *31*(2), 215–226. https://doi.org/10.1016/S0191-8869(00)00130-6

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making*, *16*(2), 77–92. https://doi.org/10.1002/bdm.433

Levin, I. P., Gaeth, G. J., Schreiber, J., & Lauriola, M. (2002). A new look at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes*, *88*(1), 411–429. https://doi.org/10.1006/obhd.2001.2983

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

Lüdecke, D., Aust, F., Crawley, S., & Ben-Shachar, M. S. (2023). *ggeffects: Create tidy data frames of marginal effects for "ggplot" from model outputs* (1.3.1) [Computer software]. https://cran.r-project.org/web/packages/ggeffects/index.html

Mandel, D. R. (2014). Do framing effects reveal irrational choice? *Journal of Experimental Psychology: General*, *143*(3), 1185–1198. https://doi.org/10.1037/a0034207

Mandel, D. R., & Kapler, I. V. (2018). Cognitive style and frame susceptibility in decision-making. *Frontiers in Psychology*, *9*, e1461. https://doi.org/10.3389/fpsyg.2018.01461

McElroy, T., Dickinson, D. L., & Levin, I. P. (2020). Thinking about decisions: An integrative approach of person and task factors. *Journal of Behavioral Decision Making*, *33*(4), 538–555. https://doi.org/10.1002/bdm.2175

McElroy, T., & Seta, J. J. (2003). Framing effects: An analytic–holistic perspective. *Journal of Experimental Social Psychology*, *39*(6), 610–617. https://doi.org/10.1016/S0022-1031(03)00036-2

Miller, P. M., & Fagley, N. S. (1991). The effects of framing, problem variations, and providing rationale on choice. *Personality and Social Psychology Bulletin*, *17*(5), 517–522.

Nieznański, M., & Obidziński, M. (2019). Verbatim and gist memory and individual differences in inhibition, sustained attention, and working memory capacity. *Journal of Cognitive Psychology*, *31*(1), 16–33. https://doi.org/10.1080/20445911.2019.1567517

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, *47*(4), 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team, Bivand, R., Carey, V. J., DebRoy, S., Eglen, S., Guha, R., Herbrandt, S., Lewin-Koh, N., Myatt, M., Nelson, M., Pfaff, B., Quistorff, B., Warmerdam, F., Weigand, S., & Free Software Foundation, Inc. (2023). *foreign: Read data stored by "Minitab", "S", "SAS", "SPSS", "Stata", "Systat", "Weka", "dBase", . . .* (0.8-85) [Computer software]. https://cran.r-project.org/web/packages/foreign/index.html

Rachev, N. R., Han, H., Lacko, D., Gelpí, R., Yamada, Y., & Lieberoth, A. (2021). Replicating the disease framing problem during the 2020 COVID-19 pandemic: A study of stress, worry, trust, and choice under risk. *PLOS ONE*, *16*(9), e0257151. https://doi.org/10.1371/journal.pone.0257151

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

Rachev, N. R., Peycheva, E. D., Kamburidis, J. A., & McElroy, T. (2023). *Does performance on the cognitive reflection test indicate cognitive miserliness? Evidence from response times, working memory capacity, warnings, and self-corrections* [Manuscript submitted for publication].

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164–171. https://doi.org/10.1027/1015-5759/a000123

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in Fuzzy-trace theory. *Judgment and Decision Making*, *7*(3), 332–359.

Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making*, *4*(4), 249–262. https://doi.org/10.1002/bdm.3960040403

Reyna, V. F., Chick, C. F., Corbin, J. C., & Hsia, A. N. (2014). Developmental reversals in risky decision making: Intelligence agents show larger decision biases than college students. *Psychological Science*, *25*(1), 76–84. https://doi.org/10.1177/0956797613497022

Reyna, V. F., Estrada, S. M., DeMarinis, J. A., Myers, R. M., Stanisz, J. M., & Mills, B. A. (2011). Neurobiological and memory models of risky decision making in adolescents versus young adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1125–1142. https://doi.org/10.1037/a0023943

Richmond, L. L., Burnett, L. K., Morrison, A. B., & Ball, B. H. (2021). Performance on the processing portion of complex working memory span tasks is related to working memory capacity estimates. *Behavior Research Methods*, *54*, 780–794. https://doi.org/10.3758/s13428-021-01645-y

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2023). *MASS: Support functions and datasets for venables and Ripley's MASS* (7.3-60) [Computer software]. https://cran.r-project.org/web/packages/MASS/index.html

Robinson, D., Hayes, A., Couch, S., Posit Software, PBC, Patil, I., Chiu, D., Gomez, M., Demeshev, B., Menne, D., Nutter, B., Johnston, L., Bolker, B., Briatte, F., Arnold, J., Gabry, J., Selzer, L., Simpson, G., . . . Reinhart, A. (2023). *broom: Convert statistical objects into tidy tibbles (1.0.5)* [Computer software]. https://cran.r-project.org/web/packages/broom/index.html

Rosseel, Y., Jorgensen, T. D., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., & Scharf, F. (2020). *lavaan: Latent variable analysis* (0.6-6) [Computer software]. https://CRAN.R-project.org/package=lavaan

Seedorff, M., Oleson, J., & McMurray, B. (2019). Maybe maximal: Good enough mixed models optimize power while controlling type I error. *PsyArXiv*. https://doi.org/10.31234/osf.io/xmhfr

Simon, A. F., Fagley, N. S., & Halleran, J. G. (2004). Decision framing: Moderating effects of individual differences and cognitive processing. *Journal of Behavioral Decision Making*, *17*(2), 77–93. https://doi.org/10.1002/bdm.463

Simonsohn, U. (2014). [17] *No-way Interactions. Data Colada*. http://datacolada.org/17

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

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M. S., Højsgaard, S., Fox, J., Lawrence, M. A., Mertens, U., Love, J., Lenth, R., & Christensen, R. H. B. (2023). *afex: Analysis of factorial experiments* (1.3-0) [Computer software]. https://cran.r-project.org/web/packages/afex/index.html

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3.

Smith, S. M., & Levin, I. P. (1996). Need for cognition and choice framing effects. *Journal of Behavioral Decision Making*, *9*(4), 283–290. https://doi.org/10.1002/(SICI)1099-0771(199612)9:4<283::AID-BDM241>3.0.CO;2-7

Stanovich, K. E., & West, R. F. (1998). Individual differences in framing and conjunction effects. *Thinking & Reasoning*, *4*(4), 289–317.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665. https://doi.org/10.1017/S0140525X00003435

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695. https://doi.org/10.1037/0022-3514.94.4.672

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The rationality quotient: Toward a test of rational thinking*. Cambridge, MA: MIT Press.

Starcke, K., Pawlikowski, M., Wolf, O. T., Altstötter-Gleich, C., & Brand, M. (2011). Decision-making under risk conditions is susceptible to interference by a secondary executive task. *Cognitive Processing*, *12*(2), 177–182. https://doi.org/10.1007/s10339-010-0387-3

Stefan, A. M., Gronau, Q. F., Schönbrodt, F. D., & Wagenmakers, E.-J. (2019). A tutorial on Bayes factor design analysis using an informed prior. *Behavior Research Methods*, *51*(3), 1042–1058. https://doi.org/10.3758/s13428-018-01189-8

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Svenson, O., & Benson III, L. (1993). Framing and time pressure in decision making. In Svenson, O., & Maule, A. J. (Eds.), *Time pressure and stress in human judgment and decision making* (pp. 133–144). Boston, MA: Springer. http://link.springer.com/chapter/10.1007/978-1-4757-6846-6_9

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

Takemura, K. (1994). Influence of elaboration on the framing of decision. *The Journal of Psychology*, *128*(1), 33–39. https://doi.org/10.1080/00223980.1994.9712709

Thompson, V. A., Pennycook, G., Trippas, D., & Evans, J. St. B. T. (2018). Do smart people have better intuitions? *Journal of Experimental Psychology: General*, *147*(7), 945–961. https://doi.org/10.1037/xge0000457

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*(4), S251–S278. https://doi.org/10.1086/296365

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

Urs, M., Goodmon, L. B., & Martin, J. (2019). Too much on my mind: Cognitive load, working memory capacity, and framing effects. *North American Journal of Psychology*, *21*(4), 739–768.

Wang, X. T. (1996). Framing effects: Dynamics and task domains. *Organizational Behavior and Human Decision Processes*, *68*(2), 145–157. https://doi.org/10.1006/obhd.1996.0095

Whitney, P., Rinehart, C. A., & Hinson, J. M. (2008). Framing effects under cognitive load: The role of working memory in risky decisions. *Psychonomic Bulletin & Review*, *15*(6), 1179–1184. https://doi.org/10.3758/PBR.15.6.1179

Wickham, H., & RStudio. (2019). *tidyverse: Easily install and load the "tidyverse"* (1.3.0) [Computer software]. https://CRAN.R-project.org/package=tidyverse

Wilke, C. O. (2020). *cowplot: Streamlined plot theme and plot annotations for "ggplot2"* (1.1.1) [Computer software]. https://CRAN.R-project.org/package=cowplot

Winter, B. (2019). *Statistics for linguists: An introduction using R*. New York: Routledge.

A. Appendix. Framing problems

A.1. Disease

Imagine that the USA/Canada is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

A.1.1. Gain frame

- If program A is adopted, 200 people will be saved.
- If program B is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved.

A.1.2. Loss frame

- If program A is adopted, 400 people will die.
- If program B is adopted, there is a 1/3 probability that nobody will die and a 2/3 probability that 600 people will die.

Which of the two programs would you favor?

A.2. Civil defense

Imagine a storage tank containing a very inflammable chemical begins to leak. The threat of an explosion is imminent. If nothing is done, 120 people are expected to be killed. The civil defense committee must choose between two interventions.

A.2.1. Gain frame

- If intervention A is chosen, 40 lives will be saved.
- If intervention B is chosen, there is a 1/3 probability of containing the threat with a saving of 120 lives and a 2/3 probability of saving no lives.

A.2.2. Loss frame
- If intervention A is chosen, 80 lives will be lost.
- If intervention B is chosen, there is a 1/3 probability of containing the threat with a loss of 0 lives and a 2/3 probability of losing 120 lives.

Which of the two interventions would you favor?

A.3. Cancer treatment

The National Cancer Institute has two possible treatments for a particular form of cancer, which is almost always fatal and kills approximately 300 people a year in the USA/Canada (Bulgaria). The institute must choose one of the treatments as the nationally adopted standard:

A.3.1. Gain frame
- If treatment A is adopted, of every 300 people who get this form of cancer, 100 will be saved.
- If treatment B is adopted, there is a 1/3 chance that all who get this form of cancer will be saved and a 2/3 chance that none who get this form of cancer will be saved.

A.3.2. Loss frame
- If treatment A is adopted, of every 300 people who get this form of cancer, 200 will die.
- If treatment B is adopted, there is a 1/3 chance that none who get this form of cancer will die and a 2/3 chance that all who get this form of cancer will die.

Which of the two treatments would you favor?

A.4. African village

(Berthet, 2021, adapted from Svenson & Benson, 1993)

Imagine an African village in which the children have been severely food poisoned. If nothing is done, 120 children are estimated to die. There are two alternative programs for curing the children:

A.4.1. Gain frame
- Program A will save 30 children.
- Program B provides a 1/4 chance that everybody is saved and a 3/4 chance that nobody is saved.

A.4.2. Loss frame
- Program A will leave 90 children to die.
- Program B provides a 1/4 chance that nobody dies and a 3/4 chance that everybody dies.

Which of the two programs would you favor?

A.5. Traffic accident

(Berthet, 2021, adapted from Wang, 1996)

Imagine that after a serious traffic accident, 100 people are stranded in a tunnel. Public authorities must choose between two interventions:

A.5.1. Gain frame
- If plan A is adopted, 25 people will be saved.
- If plan B is adopted, there is a 1/4 chance of saving all 100 people and a 3/4 chance of not saving anyone.

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press

A.5.2. Loss frame

- If plan A is adopted, 75 people will die.
- If plan B is adopted, there is a 1/4 chance that no people will die and a 3/4 chance that all 100 people will die.

Which of the two plans would you favor?

A.6. Derailed train

(Berthet, 2021, adapted from Svenson & Benson, 1993)

Imagine that a train out of control is about to derail near a big rail station. If nothing is done, the accident will cause 400 deaths. Public authorities must choose between two interventions:

A.6.1. Gain frame

- If intervention A is chosen, 100 people will be saved.
- If intervention B is chosen, there is a 1/4 chance of saving 400 people and a 3/4 chance that no one will be saved.

A.6.2. Loss frame

- If intervention A is chosen, 300 people will die.
- If intervention B is chosen, there is a 1/4 chance that no one will die and a 3/4 chance that 400 people will die.

Which of the two interventions would you favor?

Cite this article: Bogdanov, B., Corbin, J., Dobreva, S., McElroy, T., and Rachev, N. R. (2023). Working memory capacity and the risky-choice framing effect: A preregistered replication and extension of Corbin et al. (2010). *Judgment and Decision Making*, e39. https://doi.org/10.1017/jdm.2023.40

https://doi.org/10.1017/jdm.2023.40 Published online by Cambridge University Press