Journal of Experimental Psychology: Learning, Memory, and Cognition

High in Numeracy, High in Reflection, but Still Irrationally Biased: How Gist Explains Risky Choices

Valerie F. Reyna, Jordan E. Roue, Sarah M. Edelson, Aadya Singh, and M. G. Fennema

Online First Publication, February 13, 2025. https://dx.doi.org/10.1037/xlm0001441

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001441

High in Numeracy, High in Reflection, but Still Irrationally Biased: How Gist Explains Risky Choices

Valerie F. Reyna[1], Jordan E. Roue[1], Sarah M. Edelson[1], Aadya Singh[1], and M. G. Fennema[2]
[1] Department of Psychology, Cornell University
[2] Department of Accounting, Florida State University

Framing effects (risk preferences reverse for gains vs. losses) and the Allais paradox (risk preferences reverse when an option is certain vs. not) are major violations of rational choice theory. In contrast to typical samples, certified public accountants who are competent in working with probabilities and expected values should be an ideal test case for rational choice, especially high scorers on the cognitive reflection test (CRT). Although dual-process theories emphasize numeracy and cognitive reflection, fuzzy-trace theory emphasizes gist-based intuition to explain these effects among cognitively advanced decision-makers. Thus, we recruited a high-numeracy sample of certified public accountants ($N = 259$) and students ($N = 648$). We administered classic dread-disease framing, business framing, and Allais paradox problems and the CRT. Each participant received a gain and loss framing problem from different domains (one disease and one business), with presentation order counterbalanced across participants. Order of Allais problems was counterbalanced within participants. Within-participants (cross-domain) framing, between-participants (within-domain) framing, and the Allais paradox were observed for both samples. Accountants did not show domain-specific attenuation (differentially smaller framing) for business problems. Despite large expected-value differences between Allais problem options, accountants' choices resembled students' choices. Contrary to dual-process theories, CRT scores were positively related to framing for students (more framing with higher CRT) and inconsistently related for accountants, but high scorers had robust framing effects; high scorers also showed the Allais paradox. Results are consistent with fuzzy-trace theory's expectation that experts show framing effects because they rely primarily on gist-based intuition, not because they lack numeracy or cognitive reflection.

Keywords: framing effect, Allais paradox, cognitive reflection test, fuzzy-trace theory, accountants

Supplemental materials: https://doi.org/10.1037/xlm0001441.supp

Risky decision making is important in many domains of everyday life with potentially life-altering consequences. As examples, choices about whether to vaccinate for COVID-19, to have unprotected sex, to invest in the stock market, to play concussion-prone sports, to accept a plea bargain, or to screen for cancer each involve making decisions about risk (Garavito et al., 2021; Klein et al., 2023; Reyna, Broniatowski, & Edelson, 2021; Reyna & Mills, 2014; Thaler & Sunstein, 2008; Zottoli et al., 2023). Risky decision making has also been the object of much theorizing in psychology, neuroscience, and economics (Birnbaum, 2011; Glimcher & Tymula, 2023; Kahneman & Tversky, 1979; Kühberger & Tanner, 2010). Two empirical phenomena have been among the pillars of theoretical progress in decision theories: the framing effect and the Allais paradox (Allais, 1953; Kahneman, 2011). Here, we investigate both phenomena in populations that allow testing of major theoretical predictions. Furthermore, we examine how individual differences in cognitive

Andrew L. Cohen served as action editor.
Valerie F. Reyna https://orcid.org/0000-0003-2851-5797
Data are available in the Supplemental Material. Preparation of this article was supported in part by awards to Valerie F. Reyna from the National Science Foundation (Grant SES-2029420); the National Institute of Standards and Technology (Grant 60NANB22D052); the U.S. Department of Agriculture, National Institute of Food and Agriculture (Hatch/Smith-Lever 7005426); and the Institute for Trustworthy AI in Law and Society (supported by both National Science Foundation and National Institute of Standards and Technology Grant IIS-2229885). The authors acknowledge the late Robin Hogarth who began his professional life as a "chartered" accountant in the United Kingdom and whose concept of educated intuition influenced them.

Valerie F. Reyna played a lead role in conceptualization, funding acquisition, methodology, and writing–original draft and an equal role in project administration, resources, supervision, and writing–review and editing. Jordan E. Roue played a lead role in formal analysis, a supporting role in validation, and an equal role in visualization and writing–review and editing. Sarah M. Edelson played a lead role in data curation, a supporting role in formal analysis and visualization, and an equal role in investigation, supervision, validation, and writing–review and editing. Aadya Singh played a supporting role in formal analysis and an equal role in validation, visualization, and writing–review and editing. M. G. Fennema played a supporting role in conceptualization and methodology and an equal role in investigation, project administration, resources, and writing–review and editing.

Correspondence concerning this article should be addressed to Valerie F. Reyna, Department of Psychology, Cornell University, G331 Martha Van Rensselaer Hall, Ithaca, NY 14853, United States. Email: vr53@cornell.edu

reflection, a construct of dual-process theory, relate to whether decision-makers exhibit framing effects or the Allais paradox.[1]

Below, we briefly introduce the theoretical motivations for the research and, consequently, the predictions of each theoretical perspective for framing effects and the Allais paradox. In particular, we explain why accountants were selected as a focus of this study and the role this group of experts has played in prior research on numeracy (how individuals understand and use numbers) and decision making.

Theories and Predictions

Framing Effects

In a typical risky choice framing problem, decision-makers are offered a choice between a sure option and a risky gamble of equal expected value. For example, assuming 600 people are expected to die from a dread disease, a choice between saving 200 people for sure might be pitted against a one-third probability of saving 600 people and a two-third probability of saving no one. In this gain version, most respondents choose the sure option. However, these options can be framed as losses: 400 people dying for sure versus a two-third probability that 600 die and a one-third probability that none die. In the loss version, most respondents now choose the risky option. This shift in risk preference is called a "framing" effect because the same consequences (e.g., 600 die–200 saved = 400 die) elicit opposite preferences depending on how outcomes are framed (i.e., described). This dread-disease problem has been studied extensively (for reviews, see Broniatowski & Reyna, 2018; Kühberger, 1998), and the framing effect has been replicated robustly (for a meta-analysis, see Steiger & Kühberger, 2018).

Note that the expected values of the sure and risky options are equivalent: 200 saved equals one third multiplied by 600 saved and 400 die equals two thirds multiplied by 600 die. In an expected-value approach, all quantities are processed linearly, that is, using their objective numerical values. Choosing solely based on expected value would produce equal preferences between sure and risky options in a forced-choice task (i.e., indifference per the calculation above). However, as captured in expected utility theory, decision-makers are usually not indifferent; they have risk preferences. Tversky and Kahneman (1981) introduced these kinds of problems to illustrate critical tests of divergent predictions of expected utility theory versus prospect theory, still both widely used theories today (e.g., Barberis, 2013; Tymula et al., 2023). Prospect theory predicted gain–loss differences in risk preference, which was thought to rule out expected utility theory in its classic form.

Allais Paradox

The Allais (1953) paradox predated prospect theory and was cited as one of its influences (Kahneman & Tversky, 1979). The paradox also challenges expected utility theory as follows: Imagine that you have a choice between gaining $1 million for sure and a gamble composed of a 10% chance of gaining $5 million, 89% chance of gaining $1 million, and a 1% chance of gaining nothing. Many decision-makers are risk averse and thus prefer the sure option. However, imagine instead that you have a choice between two gambles: 11% chance of gaining $1 million (otherwise nothing) versus 10% chance of gaining $5 million (otherwise nothing). Now, most decision-makers become risk seeking, preferring the riskier $5 million over the $1 million.

Allais (1953) contended that either choosing the less risky option for both problems or choosing the more risky option for both problems was defensible rationally, but the overall pattern of risk aversion and risk seeking across problems violated basic axioms of rationality underpinning expected utility theory. As Kahneman (2011, p. 263) concluded: "If these were your preferences, you have committed a logical sin and violated the rules of rational choice." Similarly, Tversky and Kahneman (1981, 1986) argued that framing effects violate basic axioms of rationality because risk preferences vary for the same consequences. Some scholars suggest that the frames are not equivalent because of linguistic and pragmatic inferences (e.g., Mandel, 2023; Sher & McKenzie, 2006), but decision-makers who interpret them as equivalent show framing effects (Chick et al., 2016; Reyna, Brainerd, et al., 2021). These violations of rational choice are not just academic but have implications for business (e.g., assumptions about the rationality of markets), law (e.g., assumptions about the rationality of choices in plea bargaining), and medicine in which consistent risk preferences are assumed in eliciting patient preferences (Kahneman, 2011; Rachlinski & Wistrich, 2019; Reyna & Edelson, in press; Reyna et al., 2022; Zottoli et al., 2023).

Fuzzy-Trace Theory

Building on prospect theory, fuzzy-trace theory (FTT) goes beyond that theory and motivated the present work. FTT is relevant because, first, in head-to-head tests, FTT explains results for framing effects and the Allais paradox that alternative theories do not (for overviews, see Kühberger & Tanner, 2010; Reyna et al., 2023). For example, FTT introduced truncation effects as a test of these theories (see Reyna, 2012) encompassing a "zero effect" (Incekara-Hafalir et al., 2021; Reyna, 2021; Reyna & Brainerd, 2011); by truncating either the zero or nonzero part of the gambles, framing effects can be made to appear and disappear in accord with FTT's predictions (e.g., Broniatowski & Reyna, 2018; Chick et al., 2016; Reyna et al., 2014). Subsequent tests extended FTT's predictions to results from other tasks that are also known to violate expected utility theory and prospect theory (e.g., Payne, 2005; Reyna, Brainerd, et al., 2021; Venkatraman et al., 2014). Second, FTT is relevant to the present work because it also explains why framing effects and the Allais paradox should be observed in advanced reasoners—that is, despite high levels of numeracy, expertise, and reflection—and how gist-based and verbatim-based processing operate in parallel to produce effects.

Specifically, FTT assumes three types of mental representations are encoded when decision-makers process the options above: verbatim, ordinal gist, and categorical gist representations (see Broniatowski & Reyna, 2018; Reyna, 2023; Reyna & Brainerd, 2011; Reyna et al., 2023, for detailed processing models). Verbatim representations are symbolic representations of the literal details as presented, along with rote computations that tend to be

[1] The term "dual-process theory" has sometimes been used to include both standard dual-process approaches (e.g., Epstein, 1994; Evans & Stanovich, 2013; Kahneman, 2011) and fuzzy-trace theory, in contrast to global or unidimensional models of memory, reasoning, and decision making. Here, we contrast standard dual-process theories with fuzzy-trace theory, so we use the term "dual-process theory" in its narrower sense to refer to only the former.

performed automatically on such precise representations, including roughly computing expected value, which is observed beginning in childhood (Reyna & Brainerd, 1994, 2023; Schlottmann & Anderson, 1994). Evidence supporting encoding of expected value in adults includes critical tests in which choices shift when expected value changes (e.g., Levin et al., 2007; Reyna & Brainerd, 1995; Reyna & Brust-Renck, 2020) and in which framing manipulations (e.g., eliminating the zero arm of a gamble, such as two thirds no one saved, which subtracts nothing according to expected utility or prospect theory's predictions) often produce indifference between options when their expected values are equal (e.g., Reyna et al., 2014). According to FTT, indifference occurs, revealing the encoding of verbatim representations, when the effects of countervailing gist representations are removed by removing the zero arm of the gamble.

In addition to verbatim representations, decision-makers encode two kinds of gist: gist representations involving numerical comparisons (e.g., less–more) that capture ordinal distinctions—less precise than the linear verbatim representations of quantities—and gist representations that capture categorical distinctions that are qualitative (even less precise than ordinal gist). For the dread-disease problems, because some quantity and no quantity differ categorically, the simplest categorical distinction between outcomes boils down to a choice between saving some people versus either saving some people or saving none, favoring the sure gains option because saving some people is better than saving none (Fujita & Han, 2009; Wall et al., 2020; see also Oprea, 2024). The same categorical cleavage of outcomes creates a preference for the risky option for losses because no people dying is better than some people dying. The ordinal gist does not distinguish options in these problems because a higher chance of a lower outcome is similar overall to a lower chance of a higher outcome; ordinal gist is roughly compensatory and does not involve the zero outcome, unlike categorical gist. The verbatim representation is also not dispositive because expected values of the options are equivalent. Thus, the simplest categorical gist is predicted to undergird framing effects.

Predictions for the Allais paradox are derived using the same principles. Consider the Allais problem pitting a sure option against a gamble. Categorical gist favors the sure option over the gamble because of the possibility of gaining no money in the gamble as contrasted with some money in the sure option. Ordinal gist is again equivocal because some comparisons favor the sure option (higher probability compared with lower probability) and others favor the gamble (higher outcome in the gamble compared with lower outcome in the sure option). However, expected value in this problem is not equivalent across options but, rather, favors the gamble. Hence, although categorical gist tends to dominate preferences because of the fuzzy processing preference, verbatim processing of expected value works against that preference in this Allais problem, in contrast to the gain-frame problem (e.g., Broniatowski & Reyna, 2018; Reyna & Brust-Renck, 2020; see also Peters & Bjälkebring, 2015).

For the second problem of the Allais paradox, both options contain the categorical possibility of nothing, making the categorical gist equivocal (gain something or nothing vs. gain something or nothing; Reyna & Brainerd, 2011). Ordinal gist now favors the riskier option because $5 million is clearly "more" than $1 million, whereas 10% and 11% probability are approximately the same or similar (Leland, 1994; Rubinstein, 1988; Stevens et al., 2021; Wolfe et al., 2018). Last, expected value, as in the first Allais problem, also favors the riskier option. In the sections that follow, we discuss predictions of FTT and dual-process theory for individual differences in numeracy, expertise, and cognitive reflection.

Why Certified Public Accountants: Numeracy and Expertise

Certified public accountants (CPAs) offer special advantages in testing these hypotheses, and they are an interesting group of experts to study in their own right (Ashton, 2010; Chang et al., 2002; see also Cokely et al., 2018). First, accountants are high in numeracy. That is, they face tests of mathematical skills as students (e.g., in required coursework and in admission tests for business/accounting degrees), and their licensure examination and professional work routinely involve numerical computations. Indeed, they are taught mathematical theories of rational choice (expected value and expected utility theory) and instructed to base decisions on rational choice. Their high levels of numeracy have also been verified using validated tests of objective numeracy (Peters et al., 2019). Although some experts, such as judges, have been shown to exhibit framing effects, those groups do not necessarily score high in numeracy (Rachlinski et al., 2015). Therefore, CPAs offer an excellent population with which to test boundary conditions for alternative theories; one would expect them to be better prepared than college students in general or Mechanical Turk workers to exhibit rational choices (cf. Mandel & Kapler, 2018; Millroth et al., 2019).

In addition, CPAs would seem to have other advantages in resisting framing effects and the Allais paradox beyond numeracy, including (a) experience as professional decision-makers (students might be intelligent, but they are not experienced), (b) requirement to meet higher professional standards than regular accountants or bookkeepers (most college students in behavioral studies have not met professional standards), and (c) requirement to engage in 80 hr of continuing education every 2 years to keep their skills sharp (students lack this special training). Thus, CPAs are experts as compared with students who are novices, as usually defined in the decision-making literature (greater knowledge and experience), and experts have been argued to be less subject to cognitive biases, especially in their domain of expertise (e.g., Lueddeke & Higham, 2011; Shanteau, 1992; see also Feltz & Cokely, 2024).

To test boundary conditions for framing, we presented framing problems in the domain of expertise of accountants (a business problem; Chang et al., 2002), which might trigger training or experience that would diminish framing: domain-specific attenuation. Domain-specific attenuation would entail observing lower levels of framing in the business problem as compared with the dread-disease problem—a difference that should be larger for accountants (greater decline in framing) than for student nonexperts if domain-specific attenuation is elicited.

However, our main hypothesis is not whether CPAs would be more resistant than students but whether populations with high System 2 abilities would be resistant to cognitive biases that have been the major challenges to rational economic theory. System 2 abilities that we consider are analytical ability (specifically, numeracy) and cognitive reflection.

Many researchers have claimed that higher levels of numeracy should be associated with lower levels of cognitive biases (for reviews, see Peters, 2020; Reyna et al., 2009). In this connection, dual-process theories have identified System 1 as a source of intuitive biases with System 2 the rational, deliberative system that occasionally intervenes to censor biased responses (e.g., Epstein, 1994; Epstein et al., 1996; Kahneman, 2011; see De Neys, 2018; Evans & Stanovich, 2013, for important qualifications). Numeracy has been explicitly identified as a System 2 ability that underlies rational choice (e.g., Peters et al., 2006).

Some scholars have argued that numeracy also brings advantages in cognitive reflection (see below) and other cognitive/metacognitive skills (Sobkow et al., 2020). As examples, numerate individuals use elaborative heuristics (Cokely & Kelley, 2009), deliberate more on decision problems (Ghazal et al., 2014), are more consistent in processing probabilities (Traczyk et al., 2021), more accurately assess their judgments (Ghazal et al., 2014), search for more information (Ashby, 2017; Traczyk, Lenda, et al., 2018), and adaptively change strategies according to the structure of the decision problem (Traczyk, Sobkow, et al., 2018). Thus, the straightforward predictions based on dual-process approaches would be that accountants would be unlikely to show framing biases and would be especially unlikely to exhibit the Allais paradox given differences in expected value between options.

A fly in the ointment for these predictions is that accountants and other more numerate participants have shown *larger* biases compared with less numerate samples in some tasks (Peters et al., 2006, 2019).[2] For example, participants rate the attractiveness of a bet with a small loss (7/36 chance to win $9 and 29/36 chance to lose 5 cents) as substantially higher than the same bet without any loss (7/36 chance to win $9 and 29/36 chance to win $0; Slovic et al., 2004). This gap between rating of a loss bet and a no-loss bet was larger for the more numerate. Peters et al. (2006) argued that the highly numerate may sometimes make less "rational" responses than those lower in numeracy because they focus on numerical details and derive more "affective meaning" (feelings of goodness or badness) from numerical comparisons.

Research on FTT separates the effects of objective numeracy and ordinal numerical gist comparisons (e.g., comparisons between winning $9 and losing 5 cents or between winning $9 and winning $0), broadly corroborating the arguments above (Reyna & Brust-Renck, 2020; see also Sobkow et al., 2020). In the loss and no-loss bets, attractiveness was increased independently both by having an overall positive expected value (associated with objective numeracy) and by ordinal gist comparisons (winning $9 seems larger relative to losing 5 cents)—and these effects were different from those of categorical gist. As discussed above for framing and Allais problems, these results also illustrate how FTT's predictions incorporate effects of verbatim representations, which track expected value, ordinal gist representations, and categorical gist representations.

Thus, FTT predicts that accountants and other numerate responders would show gist-based framing and Allais effects despite their numerical computational abilities, as well as evidence of processing of verbatim expected value, reflected in increases in risky choices for Allais problem 1 compared with the gain-frame problem and increases in risky choices with increases in numeracy for both Allais problems (see Table 1; Broniatowski & Reyna, 2018; Reyna et al., 2014). These predictions hold because intuitive gist thinking is cognitively advanced in FTT—present among experts with sophisticated numerical skills (Reyna & Brainerd, 2023; Reyna & Lloyd, 2006). In fact, FTT predicted, later buttressed by meta-analyses, that framing effects emerge during the same developmental period that numerical proficiency grows (Defoe et al., 2015; Reyna & Brainerd, 1994; Reyna & Farley, 2006). Therefore, gist-based framing effects and the Allais paradox do not merely reflect superstitious thinking (Epstein, 1994; Pennycook et al., 2012; Risen, 2016) or processing of irrelevant attributes that fills in the gaps left by an absence of advanced computational abilities (Peters, 2020).

Cognitive Reflection

Like dual-process approaches, FTT also includes parameters for reflective processing, in other words, the metacognitive tendency to monitor and inhibit inconsistent or incorrect responses, which is distinct from verbatim-based and gist-based cognitive processes (e.g., Broniatowski & Reyna, 2018). However, predictions for framing depend on whether the gain–loss manipulation is within or between participants (see also LeBoeuf & Shafir, 2003; Stanovich & West, 2008). If participants receive both forms of the same problem, they are more likely to notice (by subtracting outcomes) that the gain and loss problems are numerically equivalent (Sinayev & Peters, 2015), but their responses are different, thereby inducing censoring to make their responses consistent (Chaxel & Russo, 2014). If such censoring occurs, specific order effects will be observed: Because problems cannot be compared until the second problem is presented, we assessed framing effects and the Allais paradox for second-presented versus first-presented problems to detect metacognitive censoring. This kind of censoring would be reflected in smaller framing differences and greater expected-value preference for the riskier option in Allais problems for problems presented second.

In dual-process approaches, the cognitive reflection test (CRT) measures the tendency to invoke System 2 processing to censor and inhibit biased responses, such as framing effects and the Allais paradox (Frederick, 2005; Toplak et al., 2011). The CRT is "the most popular and extensively investigated measure of individual differences in rationality" (Sobkow et al., 2020; see also Toplak et al., 2011, 2014). The test is assumed to capture whether people are able to inhibit the first ("intuitive") incorrect response and follow a correct ("reflective") solution. This thinking disposition is usually measured using puzzles, such as "a baseball bat and a ball cost $1.10 together, and the bat costs $1.00 more than the ball, how much does the ball cost?" (correct reflective answer: 5 cents; intuitive answer: 10 cents).

Cognitive reflection has been used to negatively predict poor Bayesian reasoning (Sirota & Juanchich, 2011), paranormal and religious beliefs (Pennycook et al., 2017; Shenhav et al., 2012; Sirota & Juanchich, 2018), and "gut feelings" or cognitive biases thought to indicate lack of analytical thinking (Sirota & Juanchich, 2018; Teovanović et al., 2015; Toplak et al., 2011, 2014, 2017). Higher cognitive reflection has also been associated with what decision theorists traditionally view as superior decision making (Sinayev & Peters, 2015) as well as with self-reported negative decision outcomes in real life (Juanchich et al., 2016; Sobkow et al., 2020).

Thus, to determine whether cognitive reflection (or the lack thereof) explains framing biases and the Allais paradox, we also

[2] "Fly in the ointment" is an idiomatic expression in English that refers to a drawback that was not at first apparent.

Table 1
Summary of Predictions and Evidence for Different Theoretical Processes

I. Verbatim processing of expected value (EV)
   A. For both groups (students and accountants), choices of the risky option in the Allais problems went up as CRT increased (main effect of CRT, main effect of problem, and no interaction between problem and CRT; verbatim differences in EV are identical across Allais problems).
      a. FTT predicts these results because (a) verbatim processing of EV occurs in parallel with processing of categorical gist and ordinal gist, (b) EV favors the risky option in both Allais problems, and (c) increases in numeracy and reflection produce more verbatim-based responses when EV differs.
   B. For both groups (students and accountants), choices of the risky option were higher in Allais Problem 1 than in the gain-frame problem.
      a. FTT predicts these results because (a) categorical gist supports the sure choices in both problems, (b) ordinal gist supports indifference in both problems, and (c) only the EV representation differs across these problems, favoring the risky option for Allais Problem 1.

II. Gist processing dominates despite verbatim processing (in parallel) and in advanced reasoners.
   A. In tests of prospect and other theories, FTT accounts for results that alternative theories do not, including differences between risky choice framing effects and the Allais paradox, variations such as truncation effects (deleting parts of the gamble to focus processing on gist/verbatim; Kühberger & Tanner, 2010), value allocations (allocating money to different outcomes of lotteries; Reyna et al., 2023), ratings (Reyna, Brainerd, et al., 2021), and gist-based framing effects across the lifespan (Huizenga et al., 2023).
   B. Framing effects and the Allais paradox were observed here for both groups, regardless of confidence, and at every level of CRT from low to high.
      a. FTT predicts these results because gist dominates cognition despite high reflection (CRT; Broniatowski & Reyna, 2018), high numeracy (accountants are high in numeracy; Reyna & Brust-Renck, 2020), expertise (accountants are experts; Reyna et al., 2014), and high confidence (echoing highly confident gist-based responses in false memory explained by FTT; Brainerd et al., 2017; Reyna et al., 2016).

III. Metacognitive censoring of preferences: Weak effects of problem order
   A. Censoring occurs more when problems are presented within participants rather than between participants (Stanovich & West, 2008) because problems are compared, reducing inconsistent responses in later problems. Order effects were used to detect censoring by comparing later preferences with earlier preferences.
   B. For both groups, order of gain and loss problems did not affect framing, indicating that metacognitive censoring that would reduce framing was not triggered, likely because problems within participants had unrelated content (disease and business).
   C. For both groups, order interacted with problem for Allais problems; later choices moved closer to EV (and thus were slightly more consistent), although changes were sometimes unreliable.
      a. FTT and other theories predict these results because choices become more consistent when censoring is triggered, as occurred for the Allais paradox.

IV. Evidence bearing on dual-process theory, expected value, expected utility theory, prospect theory, and FTT
   A. Contrary to predictions of dual-process theory, neither framing effects (gain–loss differences) nor the Allais paradox diminished as CRT increased.
   B. Contrary to predictions of dual-process theory, highly numerate participants (accountants) did not exhibit small or nonexistent gain–loss framing differences nor a small or nonexistent Allais paradox. Instead, differences were similar to those for undergraduates in this study and to prior samples (Broniatowski & Reyna, 2018, meta-analysis; Reyna & Brust-Renck, 2020) and did not differ for familiar content (business problems).
   C. Contrary to expected value, choices were not equally indifferent to gain and loss framing problems, and choices were not equally risky for the Allais problems.
   D. When expected utility theory is interpreted as merely consistent choices (sure–sure or risky–risky), observed choices at the group and individual level violated predictions.
   E. When expected utility theory is interpreted as consistent risk-averse choices due to the shape of the utility function (sure–sure choices), observed choices at the group and individual level violated predictions.
      a. Observed framing effects and the Allais paradox violate predictions of expected utility theory.
   F. At the group level, patterns of choices resembled those expected by prospect theory except that Allais Problem 1 tended to elicit more risky choices than the gain-frame problem. At the level of the individual, differences have been predicted using dual-process assumptions as measured by the CRT, violated as indicated above.

V. Some cognitive reflection effects observed here (e.g., increasing framing effects) and elsewhere (Broniatowski & Reyna, 2018; Cokely & Kelley, 2009) suggest that reflection can amplify gist-based intuitive responding, more easily detected when it does not conflict with verbatim-based responses.

Note. For detailed discussions of expected utility theory and consistent choices, see Blavatskyy et al. (2022), Cason and Plott (2014), Fan (2002), Harman and Gonzalez (2015), and Incekara-Hafalir et al. (2021). Predictions above apply to decisions from description, not from experience; the latter requires additional assumptions to account for learning and memory effects (e.g., Harman & Gonzalez, 2015; Reyna & Farley, 2006). CRT = cognitive reflection test; FTT = fuzzy-trace theory.

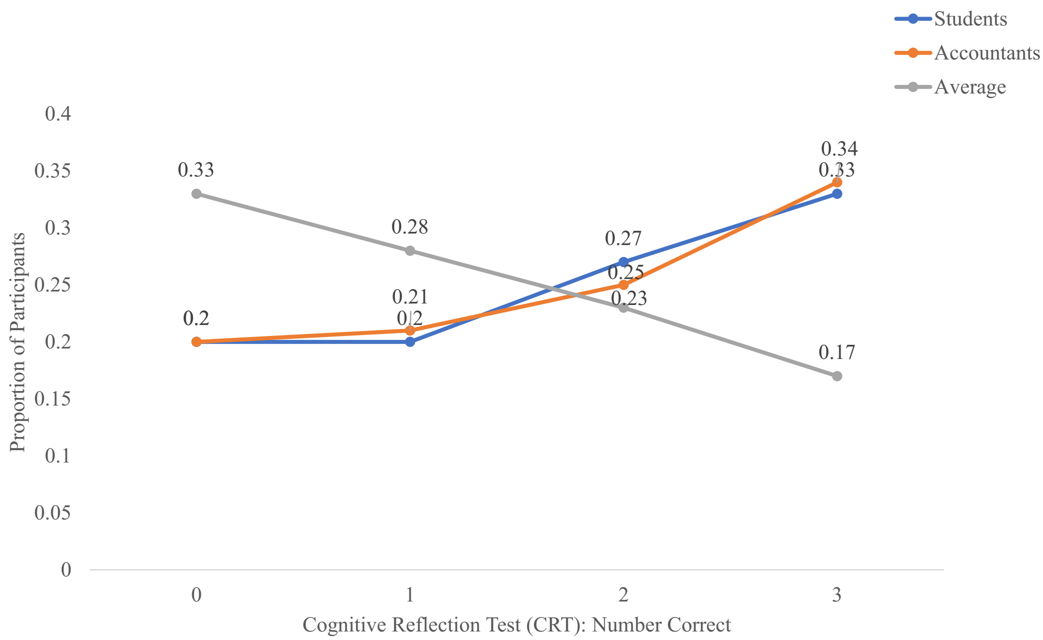
administered the CRT. The CRT remains a viable and stable measure (Bialek & Pennycook, 2018; Meyer et al., 2018; Stagnaro et al., 2018). Because the CRT likely taps numeracy as well as cognitive reflection (Liberali et al., 2012; Peters, 2020; Reyna & Brainerd, 2023), we expected that accountants would score higher than students. However, as reported below, the distributions of scores were virtually indistinguishable across accountant and student groups, facilitating comparisons of these expert professionals with students lacking professional experience and training. As shown in Figure 1, both groups scored higher on the CRT than average respondents in other studies, providing a stronger test of boundary conditions for violating rationality.

Summary

In summary, FTT predicts that framing effects and the Allais paradox will be observed despite high levels of numeracy, expertise, and reflection because the source of such effects—gist-based processing—is not the System 1 in dual-process theory. That is, gist-based responding is not the same thing as an unthinking default or

Figure 1

Proportion of Students, Accountants, and Average of 35 Other Studies



Note. Average refers to the proportions across 3,428 respondents (mostly college students) for all 35 separate studies reported by Frederick (2005) for the identical cognitive reflection test questions. See the online article for the color version of this figure.

impulsivity ("hasty, careless, inattentive" responding; cf. Oprea, 2024; cf. Frederick, 2005; Reyna et al., 2011) or lack of mental resources (cf. Evans & Stanovich, 2013; Reyna & Brainerd, 1995) or lack of understanding of the task (cf. Cason & Plott, 2014; Chick et al., 2016). Instead, gist is about distilled meaning in contrast to literal minutiae. In this study, FTT distinguishes effects of gist, verbatim processing of expected value, metacognitive monitoring and censoring, and numeracy's influence on expected-value sensitivity (see Table 1). FTT differs from dual-process theory in that it explains where framing effects and the Allais paradox come from at the level of mental representations and why they are preferred modes of responses in developmentally advanced reasoners. However, we empirically evaluate a range of hypotheses from differing perspectives, including that biases might be attenuated for accountants given their specific training and abilities, especially for domain-specific decisions (business problems) and for options that differ in expected value (Allais problems), offsetting paradoxical preferences.

Method

Participants

Two samples were recruited and included in analyses: 648 college students who participated for course credit in psychology and human development courses and 259 CPAs who participated as part of continuing professional education. The number of participants in the analyses below varies slightly because a small number of participants (4 students and 26 accountants) dropped out of the survey before completion. Results were highly similar for most and least inclusive samples; see below. (One student and 15 accountants only answered one framing question and thus were not included in any analyses.) The student sample was 69.4% female and 15.2% Hispanic (asked separately from race). Racial makeup was 46.6% White, 38.0% Asian, 10.1% Black, 1.6% American Indian, and 3.7% other. The accountant sample was 49.8% female and 21.5% Hispanic (asked separately from race). Racial makeup was 86.8% White, 6.1% Asian, 2.2% Black, and 4.8% other. The accountants had an average of 11.38 years ($SD =$ 11.48 years) of experience, ranging from 0 to 45 years, as practicing CPAs. The distribution of primary practice areas of accounting was as follows: 40.6% audit, 40.2% tax, 12.2% advisory, and 7.0% other.

Design

The design for framing problems was 2 (frame: gain or loss problem) × 2 (group: students or accountants) × 2 (problem set: disease-gain and business-loss or disease-loss and business-gain) × 2 (order: gain–loss or loss–gain), with the first factor varied within participants and the last three varied between participants. In addition to overall analyses of variance (ANOVAs) with these factors, we conducted follow-up analyses comparing framing for disease versus business domains and adding a between-participants factor of CRT number correct with four levels: 0, 1, 2, or 3. The design for Allais problems was 2 (problem: 1 or 2) × 2 (group: students or accountants) × 2 (order: Problem 1–Problem 2 or Problem 2–Problem 1), with the first factor varied within participants and the last two varied between participants. Again, we followed up adding CRT as a between-participants factor (CRT number correct).

Procedure

After providing informed consent, participants received the following instructions designed to minimize ambiguity and encourage ecological validity:

> Thank you for participating in our study. You will be making simple decisions about various scenarios. Answer like you would if you were making these decisions in real life. Treat each decision as a separate choice and answer based on how you feel at that moment.
>
> Before each decision, you will read a description of a scenario. You will then see one or two options, A and B. You will be asked to choose and to rate your preference.
>
> Please note that the probabilities and numbers presented are EXACT. For example, "$10 for sure" means that it is CERTAIN that it will be EXACTLY $10. This means that it CANNOT be $9 and it CANNOT be $11. Also, if the option specifies a probability, the chance of an event occurring will be EXACTLY the probability that is indicated.

They were also encouraged to not skip questions and to "Indicate what YOU prefer." Each participant received a gain and loss framing problem from different domains (one disease and one business), with presentation order counterbalanced across participants. (The materials are provided in the Appendix.) Next, participants responded to a scale that served as a buffer task, followed by the two Allais problems. The students received additional questions between the two Allais problems. For both students and accountants, the presentation order of Allais problems was counterbalanced across participants. After each decision, participants were asked to rate their degree of preference from 1 (*no preference: guess*) to 5 (*strongly prefer*) for framing tasks. The accountants were also asked their degree of preference for the Allais problems from 1 to 5, but for comparability to other studies, the students were asked their preference from 1 (*not at all confident*) to 7 (*completely confident*) for Allais problems. Participants then completed the CRT, which consisted of three open-ended questions as described in the Appendix and, last, demographic/background questions (e.g., race, ethnicity, and gender at birth/identification including *other* and *prefer not to answer*, years of experience and specialty for accountants and year in school and major for students).

Transparency and Openness

The research ethics committee declared this study exempt. We conducted power analyses as described in detail in Supplemental Material; results indicated that our comparisons are well powered. For all a priori comparisons, the recruited number of participants exceeded the indicated minimum sample size required, assuming .80 power, α of .05, and a medium effect size ($\eta_p^2 = 0.06$). Observed power was over .99 assuming an α of .05 to detect small effects for all main effects and interactions except four-way and five-way interactions with CRT, which were powered to detect effect sizes of $\eta_p^2 = 0.025$ or less. Even when assuming the smallest possible cell size in our design, we achieved .84 power to detect a medium effect size ($\eta_p^2 = .05$) for the five-way interaction.

We determined in advance to include all accountants who attended continuing education sessions. Students were recruited over a semester. Analyses were completed after recruitment. We report all data exclusions, manipulations, and all measures in the study, except for one scale and two sets of questions to be reported separately: For all groups, framing problems preceded any other tasks. The order of gain–loss and disease-business problems was counterbalanced. Then, for all groups, the scale followed, acting as a buffer between framing and Allais problems. For students, the additional questions were presented between the Allais problems; the order of Allais problems was counterbalanced. For all groups, the CRT followed the last Allais problem. De-identified data are reported in Supplemental Material and research materials (e.g., problem wording and instructions) are reported in the article. This study's design and its analyses are tests of published theoretical mechanisms as described but were not otherwise preregistered.

Results

For each decision, choices were scored as 0 for the less risky (e.g., sure) option and 1 for the more risky option. Three parallel sets of ANOVAs were conducted for each of the designs specified above with choice, preference/confidence ratings, and signed ratings, respectively, as dependent measures. Signed ratings were obtained by multiplying ratings by 1 if participants selected the more risky option and by −1 if the less risky option was selected, yielding a scale ranging from +5 for the strongest preference for the more risky option to −5 for the strongest preference for the less risky option (the latter being the sure option for framing problems). We rescaled the students' Allais ratings for comparability to the framing problems below by dividing each individual's rating by 7 and then multiplying it by 5.

For the bat and ball problem on the CRT, responses of "5" and ".05" were counted as correct. For the widget problem, responses of "5" were counted as correct, and for the lily pads problem, responses of "47" were counted as correct. The CRT Sum was calculated as the total number of correct responses on the bat and ball, widget, and lily pad problems for participants who gave an answer to all of the CRT problems (644 students and 233 accountants).

We focus on significant effects below but complete results, including nonsignificant effects, for all analyses are reported in Supplemental Material. An advance organizer summarizing important results can be found in Table 1.

Framing Problems

Order Effects

To begin, we examined the order of problems (gain first vs. gain second), specifically, gain–loss framing differences across order for confidence, choices, and signed confidence. There were no significant two-way interactions of frame by order, $F(1, 899) = 1.430$, $p = .232$ for confidence; $F(1, 899) = 0.280$, $p = .597$ for choices; and $F(1, 899) = 0.693$, $p = .405$ for signed confidence. If anything, for choices and signed confidence, there was a larger framing effect when problems came second rather than first, though differences occurred for the loss problems for students and for the gain problems for accountants, that is, frame, order, and group interacted, $F(1, 899) = 6.931$, $p = .009$ for choice; $F(1, 899) = 4.251$, $p = .040$ for signed confidence; see Supplemental Material. Thus, within-participant censoring of framing, such that later choices were less likely to exhibit framing than earlier choices, was not observed across problems overall nor for the same problems across earlier versus later positions. In other words, presenting choices from different domains—disease and business—was sufficient to obscure the manipulation of framing within participants that can trigger censoring preferences. Furthermore, for both students and accountants, frame and problem set interacted with gain–loss order such that business problems elicited slightly *larger* framing effects when they followed classic dread-disease problems, $F(1, 899) = 9.937$, $p = .002$. Paired comparisons showed that framing differences were nevertheless significant for every combination of problem set and order (see Supplemental Material). Thus, the framing effects discussed below occurred for both disease and business problems regardless of order.

Confidence Ratings

Decision-makers were moderately confident in their choices generally exceeding 3 on a 1–5 scale on average, clearly not indifferent or guessing (a rating of 1). In ANOVAs of frame, group, problem set, and order, there was a main effect of group, $F(1, 899) = 13.727$, $p < .001$: Accountants were significantly more confident than students ($M = 3.328$, $SE = .054$ for accountants; $M = 3.092$, $SE = .034$ for students). However, frame interacted with group, $F(1, 899) = 11.785$, $p < .001$: Accountants were significantly more confident for losses ($M = 3.420$, $SE = 0.065$) than students ($M = 3.029$, $SE = 0.041$), but the difference in the same direction for gains was not significant ($M = 3.235$, $SE = 0.067$ for accountants; $M = 3.156$, $SE = 0.042$ for students).

As noted above, the distribution of participants across levels of CRT was highly similar across groups (Figure 1). We also plotted the average proportion of 3,428 participants at each level of number correct for the same items across all 35 studies reported by Frederick (2005). Both of our samples scored high in cognitive reflection when compared against this average. Only one sample of the 35 studies scored higher. These high scores also imply high levels of numeracy (analytical ability) based on the literature (e.g., Sinayev & Peters, 2015; Weller et al., 2013; for a meta-analysis, see Otero et al., 2022).

When CRT was added as a factor to the analysis with confidence as a dependent measure, no significant effects of CRT (or interactions with CRT) were observed. However, the problem set main effect became significant, $F(1, 845) = 5.358$, $p = .021$, such that confidence for disease-gain and business-loss problems was slightly higher ($M = 3.294$, $SE = 0.048$) than that for the disease-loss and business-gain problems ($M = 3.137$, $SE = 0.048$). Group and frame by group remained significant for confidence when CRT was included as a factor (see Supplemental Material).

Preferences: Choices and Signed Confidence Ratings

Beginning with choices, in ANOVAs of frame, group, problem set, and order, there were significant main effects of frame, $F(1, 899) = 215.419$, $p < .001$, group, $F(1, 899) = 31.311$, $p < .001$, and problem set, $F(1, 899) = 16.033$, $p < .001$. That is, losses elicited more risky choices than gains overall, $M = 0.714$, $SE = 0.017$ versus $M = 0.360$, $SE = 0.017$, respectively; accountants made more risky choices than students overall, $M = 0.604$, $SE = 0.020$ versus $M = 0.471$, $SE = 0.013$, respectively; and problem sets with the disease problem as losses elicited more risky choices than the other problem set. The latter reflects a domain effect discussed below that replicated across students and accountants, namely, that the disease problems elicited larger framing effects than the business problems. Notably, frame and group did not interact with each other, $F(1, 899) = 0.036$, $p = .849$, or with problem set (i.e., the three-way interaction of frame, group, and problem set was not significant), $F(1, 899) = 0.025$, $p = .874$. As shown in Figure 2, although risky choices were translated upward for accountants compared with students, the differences between gains and losses for each set of problems were remarkably similar.[3]

When CRT was added to the analysis for choices, the Frame × CRT interaction missed significance, $F(3, 845) = 2.452$, $p = .062$, but CRT interacted with frame and group, $F(3, 845) = 3.672$, $p = .012$. Gain–loss differences at each level of CRT (Panel A) and within gains and losses across levels of CRT (Panel B) are shown in Figure 3 for students and in Figure 4 for accountants. The pattern for students was a diverging fan: As the number of problems correctly solved on the CRT increased, the framing effect increased (Figure 3, Panel B).[4] That is, risky choices increased for losses but decreased for gains with higher levels of cognitive reflection. The pattern for accountants stayed roughly the same as the number of problems correctly solved on the CRT increased, with the exception of larger framing at the second highest CRT score (Figure 4, Panel B). Although the largest framing difference for accountants was observed at the next-to-highest level of CRT performance (two correct out of three), all of the remaining differences were similar across levels of cognitive reflection. For both students and accountants, framing differences were significant at every level of CRT performance from lowest to highest (Panel A in Figures 3 and 4).

Turning to signed confidence analyses, all of the same main effects as for choices were significant, showing similar patterns of findings: frame, $F(1, 899) = 233.899$, $p < .001$; group, $F(1, 899) = 30.469$, $p < .001$; and problem set, $F(1, 899) = 15.281$, $p < .001$. As with choices, frame did not interact with group, $F(1, 899) = 0.537$, $p = .464$. When CRT was added as a factor, the Frame × CRT interaction was significant, $F(3, 845) = 3.021$, $p = .029$, as was Frame × Group × CRT, $F(3, 845) = 3.044$, $p = .028$. That is, for students, signed ratings generally increased for losses and steadily decreased for gains as CRT increased, but ratings varied nonmonotonically for accountants (see Supplemental Material for means, standard deviations, and graphs).

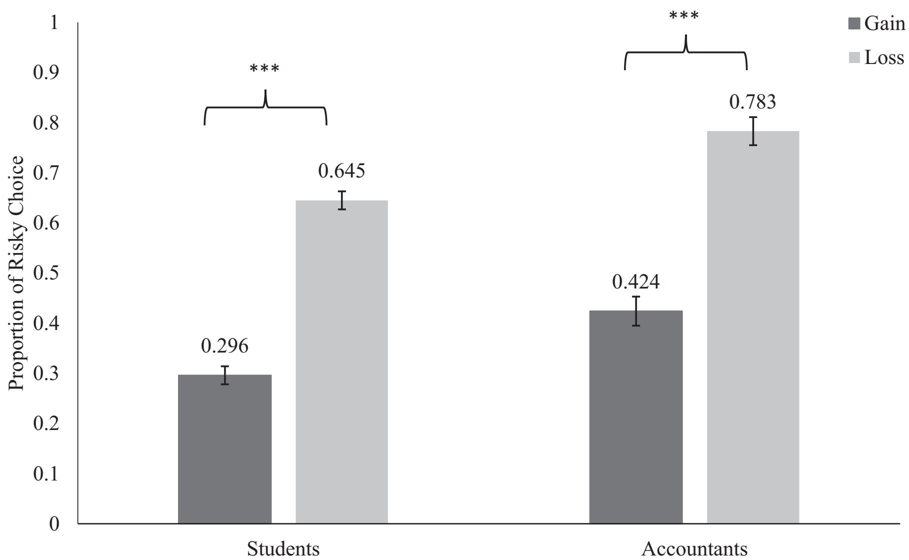
Domain Differences for Choices and Signed Confidence Ratings

To isolate domain-specific effects, we compared framing effects for disease problems with framing effects for business problems

[3] "Translated" is a mathematical term that refers to moving a shape a certain number of units (in our example, upward) without rotation or reflection about an axis.

[4] A "diverging fan" is a descriptive expression that refers to whether an interaction involves values that fan out (spread out in opposite directions), as opposed to a converging pattern or a crossover interaction.

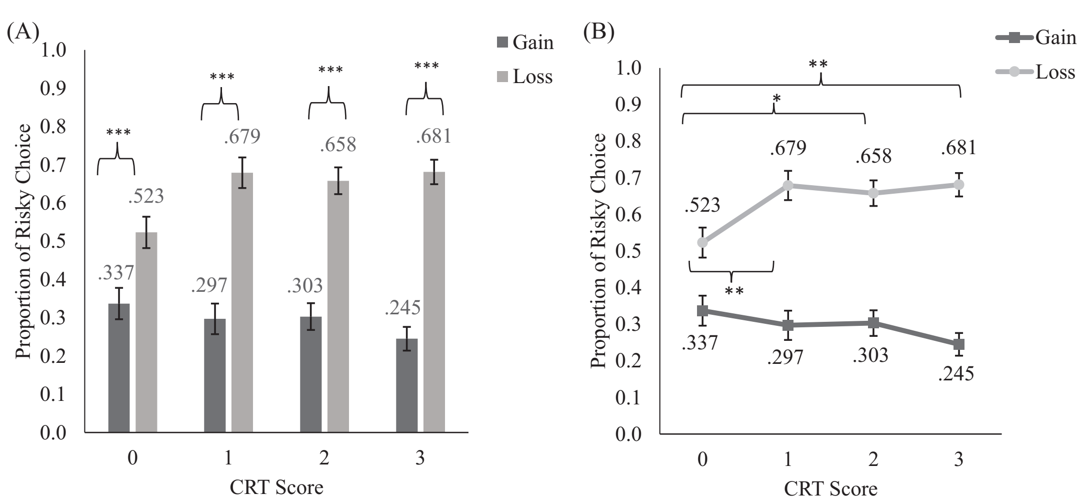
Figure 2
Framing Effects for Students and Accountants



*** $p < .001$.

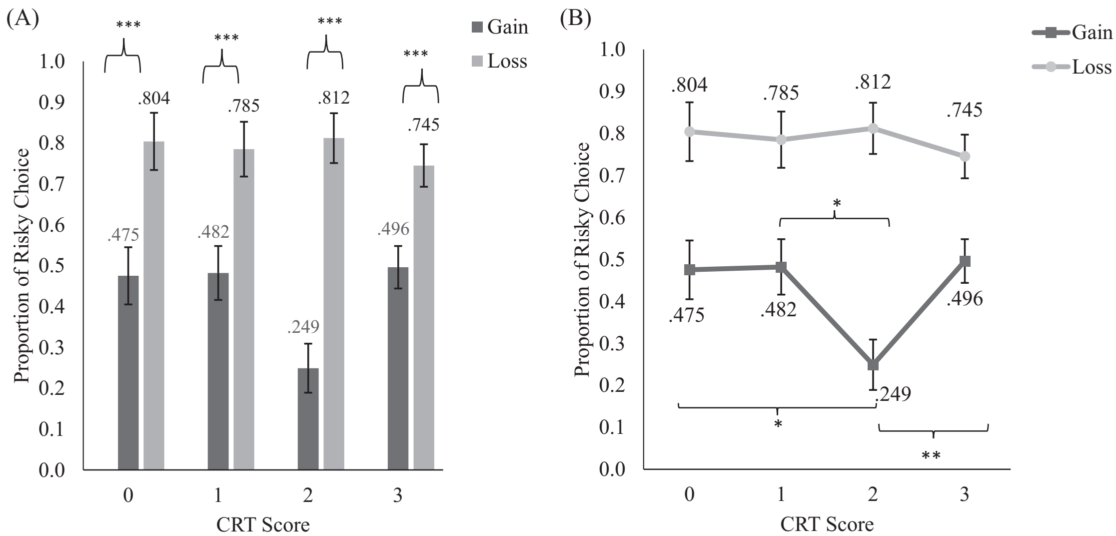
(i.e., pairwise comparisons of frame taking into account problem set) and, in subsequent analyses, added CRT. Thus, interactions between frame and problem set signal domain-specific effects. We observed a significant interaction among frame, problem set, and problem order in choice and signed confidence analyses: $F(1, 899) = 9.937$, $p = .002$ and $F(1, 899) = 6.818$, $p = .009$, respectively. That is, disease problems consistently elicited larger framing effects than business problems, but framing effects were larger for business problems when they followed disease problems (compared with when business problems came first). These effects were robust to differences in group; they were similar for students versus accountants. In addition, we observed an interaction among frame, problem set, and CRT for signed confidence ratings, $F(3, 845) = 2.698$, $p = .045$; see Supplemental Material): All within-participants pairwise comparisons of gain and loss frames were significant at every level of CRT for both problem sets. Domain differences between disease-gain and business-gain problems were significant at CRT levels of 0 and 1, but differences between disease-loss and business-loss problems were significant at CRT level 3, yielding no particular pattern across CRT. Overall, framing effects tended to be slightly smaller for business than disease problems across students and accountants, sometimes differing across domains for those low in cognitive reflection and sometimes for those high in cognitive reflection.

Figure 3
Framing by CRT Score for Students



Note. Panel A presents statistical comparisons of gains versus losses. Panel B presents statistical comparisons across groups with different CRT scores. CRT = cognitive reflection test.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 4
Framing by CRT Score for Accountants



Note. Panel A presents statistical comparisons of gains versus losses. Panel B presents statistical comparisons across groups with different CRT scores. CRT = cognitive reflection test.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Allais Problems

Order Effects

Unlike framing problems presented within participants, Allais problems had related content. In other words, order effects that trigger censoring of inconsistent responses obviously require that decision-makers realize that the content of one problem is related to the content of another problem when both problems are presented to the same person (i.e., within participants). Gain and loss framing problems were presented within participants (which sometimes triggers censoring, as discussed), but the content of the business problem was not related to the content of the disease problem, which effectively disguised that loss problems were equivalent to gain problems. The Allais problems were more similar to each other than business, and disease problems were to each other and thus might have triggered censoring and other order effects. We can analyze any effect of order because the Allais problems were counterbalanced.

Order had no effect on confidence ratings in the overall analysis, but it interacted with Allais problem when CRT was included as a factor: Confidence was slightly higher for Problem 1 when it was presented second rather than first, $F(1, 861) = 4.191$, $p = .041$. In choice and signed ratings analyses, order interacted with problem such that preference for the risky option increased when Problem 1 was presented second (after the two-gamble problem) rather than first, $F(1, 886) = 4.965$, $p = .026$ for choice and $F(1, 886) = 5.707$, $p = .017$ for signed ratings. For example, the proportion of risky choices for Problem 1 increased from .51 to .58 when the two-gamble problem that highlighted quantitative differences between options (Problem 2) appeared first; this order effect was similar for students and accountants. Although the omnibus interaction between order and problem was significant for choices, the means for Problem 1 did not differ significantly across orders in pairwise comparisons; this difference was significant for signed confidence, again, with a higher risk preference for Problem 1 when it came second. Nonetheless, as discussed below, the Allais paradox (significant differences between Problem 1 and 2) was observed for both orders of problem presentation.

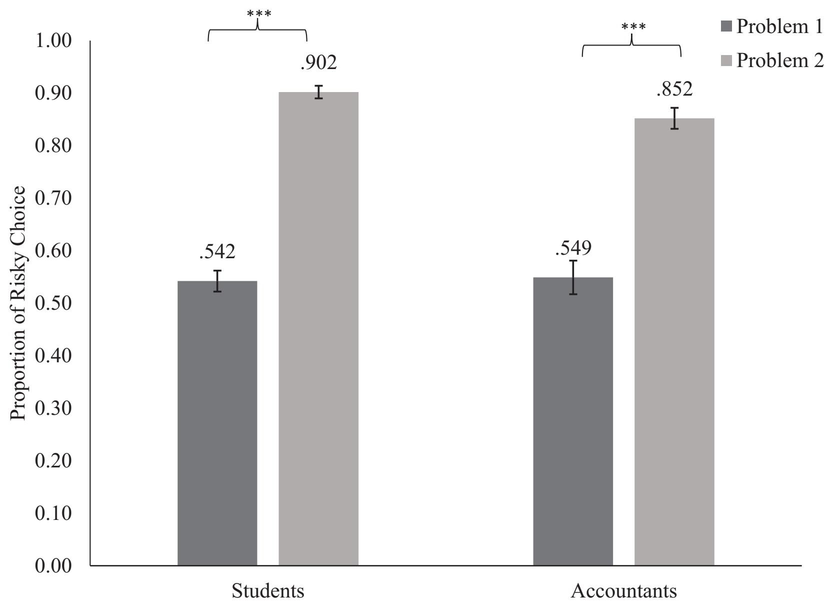
Confidence Ratings

None of the factors was significant in the overall analysis of confidence ratings. Students and accountants were both confident in their preferences for the Allais problems: a mean of 3.98 for each group (with *SE*s of 0.031 and 0.051, respectively). However, when CRT was added, that factor was significant, and it interacted with group; main effect of CRT: $F(3, 861) = 17.560$, $p < .001$ and interaction of CRT with group: $F(3, 861) = 3.127$, $p = .025$. Overall, confidence increased as CRT score increased: The lowest scoring group had lower confidence than either of the intermediate groups (which did not differ from one another) and lower confidence than the highest scoring group (the latter also differed from the second highest scoring group). Comparing students with accountants at each CRT score, confidence did not differ for groups scoring 0, 1, or 2 and only diverged at a score of 3 at which accountants were significantly more confident than students ($M = 4.32$, $SE = 0.087$ and $M = 4.11$, $SE = 0.052$, respectively). Although accountants were more confident than students, they were not less prone to the Allais paradox, as discussed below.

Preferences: Choices and Signed Confidence Ratings

As shown in Figure 5 for choices, a main effect of problem was observed for both students and accountants, $F(1, 886) = 241.691$, $p < .001$: The proportion of risky choices increased by .332 overall in Problem 2 compared with Problem 1 (.360 for the students

Figure 5
Allais Paradox for Students and Accountants



Note. Problem 1 has sure and gamble options; Problem 2 has two gambles.
*** $p < .001$.

and .303 for the accountants, a nonsignificant difference between groups), $F(1, 886) = 1.806$, $p = .179$. Signed confidence ratings also demonstrated this effect of problem, $F(1, 886) = 266.199$, $p < .001$, again with no interaction with group, $F(1, 886) = 1.943$, $p = .164$.

When CRT was added to the analyses, there was a significant main effect of CRT for both choices, $F(3, 861) = 5.474$, $p < .001$ and signed confidence, $F(3, 861) = 7.823$, $p < .001$. The preference for risk for those who scored highest in CRT (3) differed significantly from those scoring at every other level of CRT (0, 1, and 2). In other words, as CRT increased, participants were generally more likely to prefer the riskier but higher expected-value option. For both choices and signed confidence, there was no significant interaction between CRT and problem, $F(3, 861) = 0.234$, $p = .873$ and $F(3, 861) = 0.783$, $p = .503$, respectively or among CRT, problem, and group, $F(3, 861) = 1.766$, $p = .152$ and $F(3, 861) = 1.503$, $p = .212$, respectively. Figure 6 (students) and Figure 7 (accountants) display the differences between Allais problems, demonstrating the paradox, at each level of CRT. Paradoxical choices were exhibited regardless of cognitive reflection for both groups.

Comparing Gain Frame to Allais Problem 1 to Detect Verbatim Processing

As discussed, categorical gist and ordinal gist representations of gain-frame and Allais Problem 1 options are similar, but the Allais options differ in verbatim expected value, favoring the risky option. To detect the influence of verbatim processing, which should occur in parallel with gist processing, we conducted 2 (group: accountants vs. students) × 2 (task: gain problem or Allais Problem 1) ANOVAs for choice and signed confidence, respectively. As predicted, risky choices were significantly more likely to be elicited for the Allais problem than for the gain-frame problem: $F(1, 888) = 54.792$, $p < .001$ for choice and $F(1, 888) = 54.720$, $p < .001$for signed confidence. These differences between problems were significant for each of the groups in pairwise comparisons, all ps $< .005$.

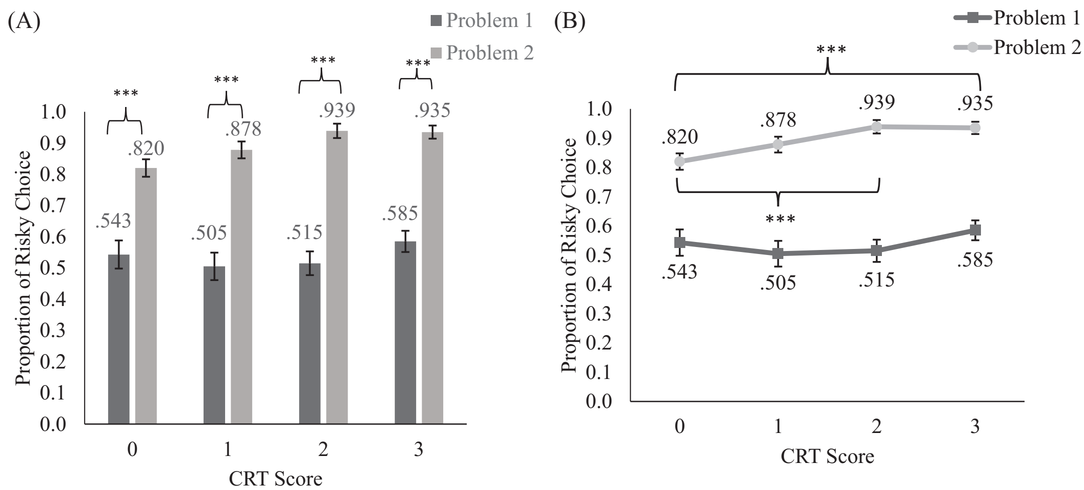
Tests of Rational Choice at the Individual Level for Framing and Allais Problems

Although framing and Allais paradox results violate expected value and expected utility theory at the group level, one might hold out hope that choices are consistent at the individual level, satisfying weak rationality (e.g., Reyna, Brainerd, et al., 2021; Wakker, 2010). Beginning with framing problems, 49% of accountants and 42% of students chose either sure–sure or risky–risky options. Taking these proportions at face value, one might argue that at least the modal pattern for each group conforms to rationality. However, this does not take error variance into account in which observed responses are subject to variability and, thus, do not directly reflect preferences; some aggregation across responses is necessary to estimate preferences, as in our group-level analyses (e.g., Levine, 1975).

We can test whether participants were indifferent between options or had consistent preferences by testing the diagonals using a chi-square goodness-of-fit test (where equal frequency is the null hypothesis; assuming consistent risk aversion, the other expected utility prediction, fails overwhelmingly). If indifference is assumed, the consistent response patterns that Cason and Plott (2014) labeled the "optimal model with noise," the diagonals should be equally likely with "random mistakes." Specifically, if deviations from consistency were just noise (i.e., randomly distributed), then the frequency of the sure(gain)–risky(loss) choices would be equivalent

Figure 6

Allais Paradox by CRT Score for Students



Note. Problem 1 has sure and gamble options; Problem 2 has two gambles. Panel A presents statistical comparisons of problem 1 versus problem 2. Panel B presents statistical comparisons across groups with different CRT scores. CRT = cognitive reflection test.

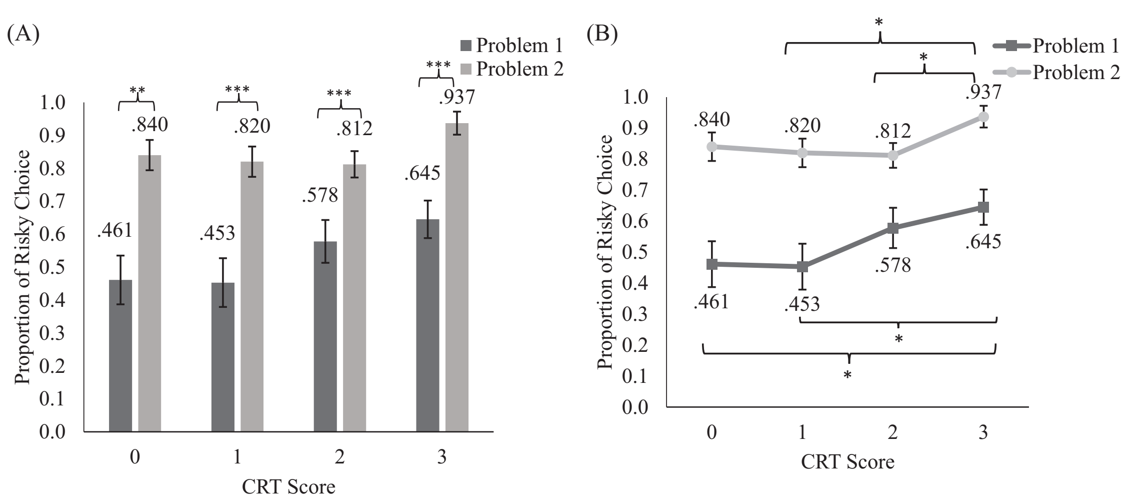
*** $p < .001$.

to the frequency of risky(gain)–sure(loss) choices. Instead, they were highly biased in favor of the former pattern, the standard framing effect: Accountants were 5.65 (113:20) times as likely to choose according to framing than the opposite, $\chi^2 = 65.03$, $p < .00001$; students were 4.01 (301:75) times as likely to choose according to framing than the opposite, $\chi^2 = 135.84$, $p < .00001$.

Turning to the Allais problem, similar results were observed at the level of individual choice patterns. Although 55% of accountants and 56% of students chose consistently (either lower risk or higher risk for both problems), accountants were 5.11 (92:18) times more likely to choose according to the Allais paradox than the opposite, and students were 9.96 (259:26) times more likely to choose according to the Allais paradox than the opposite: $\chi^2 = 49.78$, $p < .00001$ and $\chi^2 = 190.49$, $p < .00001$, respectively. For both framing problems and Allais problems, the results for the diagonals benefit from statistical aggregation and reveal that preferences were not consistent when analyzed at the level of individuals (pairs of preferences).

Figure 7

Allais Paradox by CRT Score for Accountants



Note. Problem 1 has sure and gamble options; Problem 2 has two gambles. Panel A presents statistical comparisons of problem 1 versus problem 2. Panel B presents statistical comparisons across groups with different CRT scores. CRT = cognitive reflection test.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

Rationality Violated

Rational choice theory reached its apotheosis with the axiomatization of expected utility theory (von Neumann & Morgenstern, 1944). In that view, rationality consisted of simply following basic rules of consistency in making choices. Theorems were proved showing that such a rational decision-maker would maximize expected utility, making the best choices for that individual. The first chink in the armor of rationality was the Allais paradox—shifts in risk preference when two gambles differing in riskiness were embedded in a decision with a sure option (Allais, 1953). Prospect theory inspired some of the early empirical demonstrations of the Allais paradox, hitherto mostly a thought experiment, and of framing effects (but see Edwards, 1996). The Allais paradox showed that decision-makers violate rules of rationality. Framing effects—shifts in risk preferences for the same consequences described as gains versus losses—also violate fundamental assumptions underpinning rational decision making. These violations are important because rationality continues to be an ideal in most theories of decision making, aligned with advanced System 2 analytical reasoning as compared with System 1 intuition.

A question motivating this research, one that is grounded in alternative theories of decision-making, is whether experts with purportedly requisite numerical skills and knowledge would exhibit these fundamental violations of rationality, as contrasted with inexperienced students who often populate research studies. It is not unreasonable to assume that college students might be more prone to irrationality than experienced postcollege professionals who are educated about expected utility theory. Scholars have also argued, from dual-process perspectives, that decision-makers low in numeracy and cognitive reflection should be more prone to irrationality than those higher on these dimensions. Therefore, we recruited CPAs with high levels of numeracy and with professional experience to determine whether their decisions violate rationality and compared their decisions in these classic tasks with those of students. We also used the most widely administered test of System 2 thinking—the CRT—to characterize thinkers in both groups.

We found robust violations of rationality for both students and accountants. Both groups demonstrated framing effects and the Allais paradox. Further, there was no evidence of domain-specific attenuation of violations because both groups showed slightly less and equivalent framing for the business decision compared with the disease decision. Ratings mirrored choices.

Censoring of Preferences

There was no evidence for within-participant censoring of framing effects for either group. Censoring of preferences can occur when individuals sufficiently high in System 2 thinking notice that their responses to equivalent problems are inconsistent; they change responses to make them consistent. However, we used a relatively opaque manipulation of framing within participants because gains and losses were in different domains (business vs. disease problems). When order effects were observed, they increased framing biases rather than decreased them, contrary to the censoring effect. In other words, when allowed more practice or opportunity to reflect on framing decisions (i.e., second decisions rather than first decisions), preferences became more irrational. There was some evidence of censoring for the Allais problems (a problem by order interaction), but large differences between problems remained and did not converge with CRT.

Direct Tests of Cognitive Reflection

When cognitive reflection was directly assessed, framing effects exhibited patterns inconsistent with predictions of dual-process theories. Among students, framing effects increased as levels of cognitive reflection increased; the most reflective thinkers were the most biased. Among accountants, there was nil change in framing across levels of cognitive reflection except that the largest framing biases were observed for the second highest scorers. There was nil change in the Allais paradox across levels of reflection, too (no significant interaction with problem). These results are consistent with FTT, which assumes that mental representations of verbatim quantities and of qualitative gist are independent of each other and distinct from top-down reflection or censoring. Such dissociations have been a hallmark of the theory (Reyna, 2012). Gist representations explain both framing effects and the Allais paradox (see Broniatowski & Reyna, 2018; Reyna & Brainerd, 2011; Reyna & Brust-Renck, 2020).

Confidence

Interestingly, decision-makers were confident about their choices despite violating rationality. Their ratings did not convey guessing or no preference. This is not to say that they would not be more confident for consistent choices that did not reflect conflicting mental representations (De Neys et al., 2011). However, confidence was not tantamount to less bias. Accountants were more confident than students (especially for losses), but this did not mitigate their framing effects. Those higher in CRT were more confident of their choices in the Allais problems, but this did not mitigate showing the Allais paradox.

CRT and Expected Value

CRT was associated with more risk-taking overall, regardless of the specific problem, for the Allais problems. Unlike framing problems, which do not differ in expected value, both Allais problems differ in expected value to the same degree. CRT was not observed as a main effect for the framing problems. Taken together, this pattern of results is consistent with prior findings that CRT draws on numeracy as well as reflection, which would promote choosing the riskier but higher expected-value option in both Allais problems (Bjälkebring & Peters, 2021; Liberali et al., 2012; Peters, 2020; Sobkow et al., 2020). Consistent with FTT, the Allais problem with the sure option elicited more risky choices than the gain-frame problem (Duke et al., 2018). This occurs because expected value is unequal for the Allais problems, providing a countervailing quantitative representation of the options that competes with the gist that promotes risk aversion.

Quantitative differences are highlighted for the Allais problem with two gambles because the qualitative gist representations of the option are equivalent (some money or no money for either option). Presenting that two-gamble problem first did nudge choices for the problem with the sure option toward the quantitatively superior option for both groups. These results could reflect ordinal gist comparisons rather than strictly linear expected-value comparisons in that they are reminiscent of other findings in which numerate groups were likely to choose according to relative less–more distinctions (Peters & Bjälkebring, 2015; Peters et al., 2006, 2019; Slovic et al., 2002). Nevertheless, after quantitative differences were highlighted in the first problem, a substantial Allais paradox (large differences in risk preferences across problems) remained when the problem with the sure option was subsequently presented.

Limitations and Generalizability

Limitations of this study include the use of convenience samples, which are not necessarily representative samples. However, the goals of this study were explanatory, not descriptive. Studying participants who are high in numeracy, expertise, and reflection provides a stronger test of boundary conditions on violations of rationality because they have the skills, knowledge, and cognitive propensity that theories say should allow them to resist biases—and yet they did not. Different combinations of participants and tasks might elicit different preferences (see Blavatskyy et al., 2022; Cason & Plott, 2014; Fan, 2002; Harman & Gonzalez, 2015; Incekara-Hafalir et al., 2021). These results and others raise the question of how such variability should be interpreted.

Summary

In summary, this research tests boundary conditions for exhibiting violations of rationality in decision problems that have served as pillars of the field: framing effects and the Allais paradox. By selecting a large sample of CPAs, many with years of professional experience, comparing them with students, and measuring cognitive reflection in both samples, we tested alternative predictions of expected value, expected utility theory, prospect theory, standard dual-process theory, and FTT. Our approach attempts to reconcile disparate views and findings (e.g., Mandel & Kapler, 2018; Meyer & Frederick, 2023; Millroth et al., 2019; Stanovich & Toplak, 2023; Toplak et al., 2011), distinguishing effects of gist-based intuitive risk preferences, analytical (literal) processing of expected values, cognitive reflection (sometimes positively related to intuitive biases), confidence in preferences, and metacognitive censoring of preferences. Taking a developmental approach to expertise, our results divulge an alternative conception of advanced cognition in which gist-based intuition predominates even in highly numerate and reflective reasoners.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine [Rational man's behavior in the presence of risk: Critique of the postulates and axioms of the American school]. *Econometrica*, *21*(4), 503–546. https://doi.org/10.2307/1907921

Ashby, N. J. S. (2017). Numeracy predicts preference consistency: Deliberative search heuristics increase choice consistency for choices from description and experience. *Judgment and Decision Making*, *12*(2), 128–139. https://doi.org/10.1017/S1930297500005672

Ashton, R. H. (2010). Quality, timing, and luck: Looking back at Ashton (1974). *Auditing*, *29*(1), 3–13. https://doi.org/10.2308/aud.2010.29.1.3

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *The Journal of Economic Perspectives*, *27*(1), 173–196. https://doi.org/10.1257/jep.27.1.173

Bialek, M., & Pennycook, G. (2018). The cognitive reflection test is robust to multiple exposures. *Behavior Research Methods*, *50*(5), 1953–1959. https://doi.org/10.3758/s13428-017-0963-x

Birnbaum, M. H. (2011). Testing theories of risky decision making via critical tests. *Frontiers in Psychology*, *2*, Article 315. https://doi.org/10.3389/fpsyg.2011.00315

Bjälkebring, P., & Peters, E. (2021). Money matters (especially if you are good at math): Numeracy, verbal intelligence, education, and income in satisfaction judgments. *PLOS ONE*, *16*(11), Article e0259331. https://doi.org/10.1371/journal.pone.0259331

Blavatskyy, P., Ortmann, A., & Panchenko, V. (2022). On the experimental robustness of the Allais paradox. *American Economic Journal: Microeconomics*, *14*(1), 143–163. https://doi.org/10.1257/mic.20190153

Brainerd, C. J., Nakamura, K., Reyna, V. F., & Holliday, R. E. (2017). Overdistribution illusions: Categorical judgments produce them, confidence ratings reduce them. *Journal of Experimental Psychology. General*, *146*(1), 20–40. https://doi.org/10.1037/xge0000242

Broniatowski, D. A., & Reyna, V. F. (2018). A formal model of fuzzy-trace theory: Variations on framing effects and the Allais paradox. *Decision*, *5*(4), 205–252. https://doi.org/10.1037/dec0000083

Cason, T. N., & Plott, C. R. (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy*, *122*(6), 1235–1270. https://doi.org/10.1086/677254

Chang, C., Yen, S. H., & Duh, R. (2002). An empirical examination of competing theories to explain the framing effect in accounting-related decision. *Behavioral Research in Accounting*, *14*(1), 35–64. https://doi.org/10.2308/bria.2002.14.1.35

Chaxel, A.-S., & Russo, J. E. (2014). Cognitive consistency: Cognitive and motivational perspectives. In E. A. Wilhelms & V. F. Reyna (Eds.), *Neuroeconomics, judgment, and decision making* (pp. 29–48). Psychology 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., Feltz, A., Ghazal, S., Allan, J. N., Petrova, D., & Garcia-Retamero, R. (2018). Skilled decision theory: From intelligence to numeracy and expertise. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge handbook of expertise and expert performance* (2nd ed., pp. 476–505). Cambridge University Press. https://doi.org/10.1017/9781316480748.026

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. https://doi.org/10.1017/S193029750000067X

De Neys, W. (Ed.). (2018). *Dual process theory 2.0*. Routledge/Taylor & Francis Group.

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLOS ONE*, *6*(1), Article e15954. https://doi.org/10.1371/journal.pone.0015954

Defoe, I. N., Dubas, J. S., Figner, B., & van Aken, M. A. (2015). A meta-analysis on age differences in risky decision making: Adolescents versus children and adults. *Psychological Bulletin*, *141*(1), 48–84. https://doi.org/10.1037/a0038088

Duke, K. E., Goldsmith, K., & Amir, O. (2018). Is the preference for certainty always so certain? *Journal of the Association for Consumer Research*, *3*(1), 63–80. https://doi.org/10.1086/695776

Edwards, K. D. (1996). Prospect theory: A literature review. *International Review of Financial Analysis*, *5*(1), 19–38. https://doi.org/10.1016/S1057-5219(96)90004-6

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709–724. https://doi.org/10.1037/0003-066X.49.8.709

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, *71*(2), 390–405. https://doi.org/10.1037/0022-3514.71.2.390

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Fan, C. P. (2002). Allais paradox in the small. *Journal of Economic Behavior & Organization*, *49*(3), 411–421. https://doi.org/10.1016/S0167-2681(02)00012-4

Feltz, A., & Cokely, E. T. (2024). *Diversity and disagreement: From fundamental biases to ethical interactions*. Palgrave Macmillan Cham. https://doi.org/10.1007/978-3-031-61935-9

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Fujita, K., & Han, H. A. (2009). Moving beyond deliberative control of impulses: The effect of construal levels on evaluative associations in self-control conflicts. *Psychological Science*, *20*(7), 799–804. https://doi.org/10.1111/j.1467-9280.2009.02372.x

Garavito, D. M. N., Reyna, V. F., DeTello, J. E., Landow, B. R., & Tarpinian, L. M. (2021). Intentions to report concussion symptoms in nonprofessional athletes: A fuzzy-trace theory approach. *Applied Cognitive Psychology*, *35*(1), 26–38. https://doi.org/10.1002/acp.3734

Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making*, *9*(1), 15–34. https://doi.org/10.1017/S1930297500004952

Glimcher, P. W., & Tymula, A. A. (2023). Expected subjective value theory (ESVT): A representation of decision under risk and certainty. *Journal of Economic Behavior & Organization*, *207*, 110–128. https://doi.org/10.1016/j.jebo.2022.12.013

Harman, J. L., & Gonzalez, C. (2015). Allais from experience: Choice consistency, rare events, and common consequences in repeated decisions. *Journal of Behavioral Decision Making*, *28*(4), 369–381. https://doi.org/10.1002/bdm.1855

Huizenga, H. M., Zadelaar, J. N., Jansen, B. R. J., Olthof, M. C., Steingroever, H., Dekkers, L. M. S., van Duijvenvoorde, A. C. K., Figner, B., & Agelink van Rentergem, J. (2023). Formal models of differential framing effects in decision making under risk. *Decision*, *10*(3), 197–234. https://doi.org/10.1037/dec0000201

Incekara-Hafalir, E., Kim, E., & Stecher, J. D. (2021). Is the Allais paradox due to appeal of certainty or aversion to zero? *Experimental Economics*, *24*(3), 751–771. https://doi.org/10.1007/s10683-020-09678-4

Juanchich, M., Dewberry, C., Sirota, M., & Narendran, S. (2016). Cognitive reflection predicts real-life decision outcomes, but not over and above personality and decision-making styles. *Journal of Behavioral Decision Making*, *29*(1), 52–59. https://doi.org/10.1002/bdm.1875

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

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Klein, W. M. P., Chou, W. S., & Vanderpool, R. C. (2023). Health information in 2023 (and beyond): Confronting emergent realities with health communication science. *JAMA*, *330*(12), 1131–1132. https://doi.org/10.1001/jama.2023.15817

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

Kühberger, A., & Tanner, C. (2010). Risky choice framing: Task versions and a comparison of prospect theory and fuzzy-trace theory. *Journal of Behavioral Decision Making*, *23*(3), 314–329. https://doi.org/10.1002/bdm.656

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

Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, *9*(2), 151–172. https://doi.org/10.1007/BF01064183

Levin, I. P., Weller, J. A., Pederson, A. A., & Harshman, L. A. (2007). Age-related differences in adaptive decision making: Sensitivity to expected value in risky choice. *Judgment and Decision Making*, *2*(4), 225–233. https://doi.org/10.1017/S1930297500000553

Levine, M. (1975). *Cognitive theory of learning*. Erlbaum.

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making*, *25*(4), 361–381. https://doi.org/10.1002/bdm.752

Lueddeke, S. E., & Higham, P. A. (2011). Expertise and gambling: Using type 2 signal detection theory to investigate differences between regular gamblers and nongamblers. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *64*(9), 1850–1871. https://doi.org/10.1080/17470218.2011.584631

Mandel, D. R. (2023). Subjective equivalence: A basic requirement for strict framing effects: Commentary on Huizenga et al. (2023). *Decision*, *10*(3), 235–237. https://doi.org/10.1037/dec0000206

Mandel, D. R., & Kapler, I. V. (2018). Cognitive style and frame susceptibility in decision-making. *Frontiers in Psychology*, *9*, Article 1461. https://doi.org/10.3389/fpsyg.2018.01461

Meyer, A., & Frederick, S. (2023). The formation and revision of intuitions. *Cognition*, *240*, Article 105380. https://doi.org/10.1016/j.cognition.2023.105380

Meyer, A., Zhou, E., & Frederick, S. (2018). The non-effects of repeated exposure to the cognitive reflection test. *Judgment and Decision Making*, *13*(3), 246–259. https://doi.org/10.1017/S1930297500007683

Millroth, P., Nilsson, H., & Juslin, P. (2019). The decision paradoxes motivating Prospect Theory: The prevalence of the paradoxes increases with numerical ability. *Judgment and Decision Making*, *14*(4), 513–533. https://doi.org/10.1017/S1930297500006161

Oprea, R. (2024). Decisions under risk are decisions under complexity. *The American Economic Review*, *114*(12), 3789–3811. https://doi.org/10.1257/aer.20221227

Otero, I., Salgado, J. F., & Moscoso, S. (2022). Cognitive reflection, cognitive intelligence, and cognitive abilities: A meta-analysis. *Intelligence*, *90*, Article 101614. https://doi.org/10.1016/j.intell.2021.101614

Payne, J. W. (2005). It is whether you win or lose: The importance of the overall probabilities of winning or losing in risky choice. *Journal of Risk and Uncertainty*, *30*(1), 5–19. https://doi.org/10.1007/s11166-005-5831-x

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, *123*(3), 335–346. https://doi.org/10.1016/j.cognition.2012.03.003

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin & Review*, *24*(6), 1774–1784. https://doi.org/10.3758/s13423-017-1242-7

Peters, E. (2020). *Innumeracy in the wild: Misunderstanding and misusing numbers*. Oxford University Press. https://doi.org/10.1093/oso/9780190861094.001.0001

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

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

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

Rachlinski, J., & Wistrich, A. (2019). Gains, losses, and judging: Framing the judiciary. *The Notre Dame Law Review*, *94*, 521–582. https://scholarship.law.nd.edu/ndlr/vol94/iss2/2

Rachlinski, J., Wistrich, A., & Guthrie, C. (2015). Can judges make reliable numeric judgments? Distorted damages and skewed sentences. *Indiana Law Journal*, *90*(2), 695–739. https://www.repository.law.indiana.edu/ilj/vol90/iss2/6

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in Fuzzy-Trace Theory. *Judgment and Decision Making*, *7*(3), 332–359. https://doi.org/10.1017/S1930297500002291

Reyna, V. F. (2021). A scientific theory of gist communication and misinformation resistance, with implications for health, education, and policy. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(15), Article e1912441117. https://doi.org/10.1073/pnas.1912441117

Reyna, V. F. (2023). Models of risky choice across ages, frames, and individuals: The fuzzy frontier. *Decision*, *10*(3), 238–242. https://doi.org/10.1037/dec0000209

Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: A review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 239–272). Wiley.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*(1), 1–75. https://doi.org/10.1016/1041-6080(95)90031-4

Reyna, V. F., & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review*, *31*(2–3), 180–206. https://doi.org/10.1016/j.dr.2011.07.004

Reyna, V. F., & Brainerd, C. J. (2023). Numeracy, gist, literal thinking and the value of nothing in decision making. *Nature Reviews Psychology*, *2*(7), 421–439. https://doi.org/10.1038/s44159-023-00188-7

Reyna, V. F., Brainerd, C. J., Chen, Z., & Bookbinder, S. H. (2021). Explaining risky choices with judgments: Framing, the zero effect, and the contextual relativity of gist. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(7), 1037–1053. https://doi.org/10.1037/xlm0001016

Reyna, V. F., Broniatowski, D. A., & Edelson, S. M. (2021). Viruses, vaccines, and COVID-19: Explaining and improving risky decision-making. *Journal of Applied Research in Memory and Cognition*, *10*(4), 491–509. https://doi.org/10.1016/j.jarmac.2021.08.004

Reyna, V. F., & Brust-Renck, P. G. (2020). How representations of number and numeracy predict decision paradoxes: A fuzzy-trace theory approach. *Journal of Behavioral Decision Making*, *33*(5), 606–628. https://doi.org/10.1002/bdm.2179

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., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition*, *5*(1), 1–9. https://doi.org/10.1016/j.jarmac.2015.12.003

Reyna, V. F., Edelson, S., Hayes, B., & Garavito, D. (2022). Supporting health and medical decision-making: Findings and insights from fuzzy-trace theory. *Medical Decision Making*, *42*(6), 741–754. https://doi.org/10.1177/0272989X221105473

Reyna, V. F., & Edelson, S. M. (in press). The science of risk communication: Risk-as-information, risk-as-feelings, and risk-as-gist. In L. Horton, Y. Landa, J. Schiffman, & S. Woods (Eds.), *Talking about risk for psychosis: A guide for practice and research*. Oxford University Press.

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

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. *Psychological Science in the Public Interest*, *7*(1), 1–44. https://doi.org/10.1111/j.1529-1006.2006.00026.x

Reyna, V. F., & Lloyd, F. J. (2006). Physician decision making and cardiac risk: Effects of knowledge, risk perception, risk tolerance, and fuzzy processing. *Journal of Experimental Psychology: Applied*, *12*(3), 179–195. https://doi.org/10.1037/1076-898X.12.3.179

Reyna, V. F., & Mills, B. A. (2014). Theoretically motivated interventions for reducing sexual risk taking in adolescence: A randomized controlled experiment applying fuzzy-trace theory. *Journal of Experimental Psychology: General*, *143*(4), 1627–1648. https://doi.org/10.1037/a0036717

Reyna, V. F., Müller, S. M., & Edelson, S. M. (2023). Critical tests of fuzzy trace theory in brain and behavior: Uncertainty across time, probability, and development. *Cognitive, Affective & Behavioral Neuroscience*, *23*(3), 746–772. https://doi.org/10.3758/s13415-022-01058-0

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943–973. https://doi.org/10.1037/a0017327

Risen, J. L. (2016). Believing what we do not believe: Acquiescence to superstitious beliefs and other powerful intuitions. *Psychological Review*, *123*(2), 182–207. https://doi.org/10.1037/rev0000017

Rubinstein, A. (1988). Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory*, *46*(1), 145–153. https://doi.org/10.1016/0022-0531(88)90154-8

Schlottmann, A., & Anderson, N. (1994). Children's judgments of expected value. *Developmental Psychology*, *30*(1), 56–66. https://doi.org/10.1037/0012-1649.30.1.56

Shanteau, J. (1992). The psychology of experts: An alternative view. In G. Wright & F. Bolger (Eds.), *Expertise and decision support* (pp. 11–23). Springer. https://doi.org/10.1007/978-0-585-34290-0_2

Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General*, *141*(3), 423–428. https://doi.org/10.1037/a0025391

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition*, *101*(3), 467–494. https://doi.org/10.1016/j.cognition.2005.11.001

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, *6*, Article 532. https://doi.org/10.3389/fpsyg.2015.00532

Sirota, M., & Juanchich, M. (2011). Role of numeracy and cognitive reflection in Bayesian reasoning with natural frequencies. *Studia Psychologica*, *53*(2), 151–161.

Sirota, M., & Juanchich, M. (2018). Effect of response format on cognitive reflection: Validating a two- and four-option multiple choice question version of the cognitive reflection test. *Behavior Research Methods*, *50*(6), 2511–2522. https://doi.org/10.3758/s13428-018-1029-4

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). Rational actors or rational fools: Implications of the affect heuristic for behavioral economics. *Journal of Socio-Economics*, *31*(4), 329–342. https://doi.org/10.1016/S1053-5357(02)00174-9

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, *24*(2), 311–322. https://doi.org/10.1111/j.0272-4332.2004.00433.x

Sobkow, A., Zaleskiewicz, T., Petrova, D., Garcia-Retamero, R., & Traczyk, J. (2020). Worry, risk perception, and controllability predict intentions toward COVID-19 preventive behaviors. *Frontiers in Psychology*, *11*, Article 582720. https://doi.org/10.3389/fpsyg.2020.582720

Stagnaro, M. N., Pennycook, G., & Rand, D. G. (2018). Performance on the cognitive reflection test is stable across time. *Judgment and Decision Making*, *13*(3), 260–267. https://doi.org/10.1017/S1930297500007695

Stanovich, K. E., & Toplak, M. E. (2023). Actively open-minded thinking and its measurement. *Journal of Intelligence*, *11*(2), Article 27. https://doi.org/10.3390/jintelligence11020027

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

Steiger, A., & Kühberger, A. (2018). A meta-analytic re-appraisal of the framing effect. *Zeitschrift Für Psychologie*, *226*(1), 45–55. https://doi.org/10.1027/2151-2604/a000321

Stevens, J. R., Cully, T., & Goh, F. W. (2021). *Framing attributes in similarity judgments and intertemporal choices* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/csnqv

Teovanović, P., Knežević, G., & Stankov, L. (2015). Individual differences in cognitive biases: Evidence against one-factor theory of rationality. *Intelligence*, *50*, 75–86. https://doi.org/10.1016/j.intell.2015.02.008

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275–1289. https://doi.org/10.3758/s13421-011-0104-1

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

Toplak, M. E., West, R. F., & Stanovich, K. E. (2017). Real-world correlates of performance on heuristics and biases tasks in a community sample. *Journal of Behavioral Decision Making*, *30*(2), 541–554. https://doi.org/10.1002/bdm.1973

Traczyk, J., Fulawka, K., Lenda, D., & Zaleskiewicz, T. (2021). Consistency in probability processing as a function of affective context and numeracy. *Journal of Behavioral Decision Making*, *34*(2), 228–246. https://doi.org/10.1002/bdm.2206

Traczyk, J., Lenda, D., Serek, J., Fulawka, K., Tomczak, P., Strizyk, K., Polec, A., Zjawiony, P., & Sobkow, A. (2018). Does fear increase search effort in more numerate people? An experimental study investigating information acquisition in a decision from experience task. *Frontiers in Psychology*, *9*, Article 1203. https://doi.org/10.3389/fpsyg.2018.01203

Traczyk, J., Sobkow, A., Fulawka, K., Kus, J., Petrova, D., & Garcia-Retamero, R. (2018). Numerate decision makers don't use more effortful strategies unless it pays: A process tracing investigation of skilled and adaptive strategy selection in risky decision making. *Judgment and Decision Making*, *13*(4), 372–381. https://doi.org/10.1017/S1930297500009244

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. *The Journal of Business*, *59*(4), S251–S278. https://doi.org/10.1086/296365

Tymula, A., Wang, X., Imaizumi, Y., Kawai, T., Kunimatsu, J., Matsumoto, M., & Yamada, H. (2023). Dynamic prospect theory: Two core decision theories coexist in the gambling behavior of monkeys and humans. *Science Advances*, *9*(20), Article eade7972. https://doi.org/10.1126/sciadv.ade7972

Venkatraman, V., Payne, J. W., & Huettel, S. A. (2014). An overall probability of winning heuristic for complex risky decisions: Choice and eye fixation evidence. *Organizational Behavior and Human Decision Processes*, *125*(2), 73–87. https://doi.org/10.1016/j.obhdp.2014.06.003

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511779329

Wall, D., Crookes, R. D., Johnson, E. J., & Weber, E. U. (2020). Risky choice frames shift the structure and emotional valence of internal arguments: A query theory account of the unusual disease problem. *Judgment and Decision Making*, *15*(5), 685–703. https://doi.org/10.1017/S1930297500007877

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, *26*(2), 198–212. https://doi.org/10.1002/bdm.1751

Wolfe, C. R., Reyna, V. F., & Smith, R. J. (2018). On judgments of approximately equal. *Journal of Behavioral Decision Making*, *31*(1), 151–163. https://doi.org/10.1002/bdm.2061

Zottoli, T. M., Helm, R. K., Edkins, V. A., & Bixter, M. T. (2023). Developing a model of guilty plea decision-making: Fuzzy-trace theory, gist, and categorical boundaries. *Law and Human Behavior*, *47*(3), 403–421. https://doi.org/10.1037/lhb0000532

(*Appendix follows*)

Appendix

Materials

The following tasks were administered. Stimuli are formatted as presented.

Dread Disease: Gain

Imagine the U.S. is preparing for the outbreak of an unusual dread disease, which is expected to kill 600 people. There are two options. Which would you choose?

A. 200 people saved for sure.

B. 1/3 probability 600 people saved and 2/3 probability no one saved.

Dread Disease: Loss

Imagine the U.S. is preparing for the outbreak of an unusual dread disease, which is expected to kill 600 people. There are two options. Which would you choose?

A. 400 people die for sure.

B. 2/3 probability 600 people die and 1/3 probability no one dies.

Business: Gain

Your company's current pollution control system no longer meets the minimum requirement. The company may be subject to a $300,000 punitive fine. Choose the option that you prefer:

A. The company will, for certain, be able to save $100,000 of the punitive fine.

B. There is a 1/3 probability that the $300,000 punitive fine will be saved and a 2/3 probability that a $0 punitive fine will be saved.

Business: Loss

Your company's current pollution control system no longer meets the minimum requirement. The company may be subject to a $300,000 punitive fine. Choose the option that you prefer:

A. The company will, for certain, be subject to a $200,000 punitive fine.

B. There is a 1/3 probability that the company will be subject to a $0 punitive fine and a 2/3 probability that the company will be subject to a $300,000 punitive fine.

Allais Paradox: Problem 1

Which would you choose?

A. $1 million with 1.0 probability

B. $1 million with .89 probability, nothing with .01 probability, and $5 million with .10 probability

Allais Paradox: Problem 2

Which would you choose?

A. $1 million with .11 probability and nothing with .89 probability

B. $5 million with .10 probability and nothing with .90 probability

Cognitive Reflection Test

A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost?

If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets?

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Received February 21, 2024
Revision received July 30, 2024
Accepted October 28, 2024 ■