

# Journal of Experimental Psychology: General

## Observers Penalize Decision Makers Whose Risk Preferences Are Unaffected by Loss-Gain Framing

Charles A. Dorison and Blake H. Heller

Online First Publication, February 10, 2022. <http://dx.doi.org/10.1037/xge0001187>

### CITATION

Dorison, C. A., & Heller, B. H. (2022, February 10). Observers Penalize Decision Makers Whose Risk Preferences Are Unaffected by Loss-Gain Framing. *Journal of Experimental Psychology: General*. Advance online publication. <http://dx.doi.org/10.1037/xge0001187>

# Observers Penalize Decision Makers Whose Risk Preferences Are Unaffected by Loss–Gain Framing

Charles A. Dorison<sup>1</sup> and Blake H. Heller<sup>2, 3</sup>

<sup>1</sup> Department of Management and Organizations, Kellogg School of Management, Northwestern University

<sup>2</sup> Harvard Graduate School of Education, Harvard University

<sup>3</sup> Peabody College of Education and Human Development, Vanderbilt University

A large interdisciplinary body of research on human judgment and decision making documents systematic deviations between prescriptive decision models (i.e., how individuals *should* behave) and descriptive decision models (i.e., how individuals *actually* behave). One canonical example is the loss–gain framing effect on risk preferences: the robust tendency for risk preferences to shift depending on whether outcomes are described as losses or gains. Traditionally, researchers argue that decision makers should always be immune to loss–gain framing effects. We present three preregistered experiments ( $N = 1,954$ ) that qualify this prescription. We predict and find that while third-party observers penalize decision makers who make risk-averse (vs. risk-seeking) choices when choice outcomes are framed as losses, this result reverses when outcomes are framed as gains. This reversal holds across five social perceptions, three decision contexts, two sample populations of United States adults, and with financial stakes. This pattern is driven by the fact that observers themselves fall victim to framing effects and socially derogate (and financially punish) decision makers who disagree. Given that individuals often care deeply about their reputation, our results challenge the long-standing prescription that they should always be immune to framing effects. The results extend understanding not only for decision making under risk, but also for a range of behavioral tendencies long considered irrational biases. Such understanding may ultimately reveal not only why such biases are so persistent but also novel interventions: our results suggest a necessary focus on social and organizational norms.

**Keywords:** framing effects, judgment and decision making, organizational behavior, reputation, risk taking

Drawing on research in psychology, economics, and neuroscience, more than four decades of research on human judgment and decision making document systematic deviations between prescriptive models of decision making (i.e., models describing how individuals *should* behave) and descriptive models of decision making (i.e., models describing how individuals *actually* behave; for reviews, see Arkes, 1981, 1991; Bazerman &

Moore, 2009; Fischhoff & Broomell, 2020; Gilovich & Griffin, 2010; Gilovich et al., 2002; Kahneman et al., 1991; Loewenstein et al., 2008; Loewenstein et al., 2001; Mellers et al., 1998; Payne et al., 1992; Weber & Johnson, 2009). One canonical example of these deviations is loss–gain framing effects on risk preferences: the robust tendency for risk preferences to shift depending on whether outcomes are described as losses or gains (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981). Specifically, decision makers are more likely to make risk-seeking choices when the potential outcomes of alternative choices are framed as losses and are more likely to make risk-averse choices when outcomes are framed as gains, even when choice sets are otherwise equivalent.

Researchers across disciplines have documented such framing effects on a wide array of important choices (for review, see Ruggeri et al., 2020). These include but are not limited to: buying insurance (Hershey & Schoemaker, 1980), saving for retirement (Benartzi & Thaler, 1995), reaching agreement in negotiations (Bazerman, 1983), and trading commodities (Sun & Mellers, 2016). Loss–gain framing effects on risk preferences generalize in samples ranging from professional physicians (McNeil et al., 1982) to capuchin monkeys (Chen et al., 2006). More recently, an international project successfully replicated such framing effects across 19 countries and 13 languages, concluding that these effects

---

Charles A. Dorison  <https://orcid.org/0000-0002-7072-2530>

Blake H. Heller  <https://orcid.org/0000-0003-2093-8170>

The research project was supported in part by a training grant from the Institute of Education Sciences (R305B150010). Opinions expressed do not represent the views of the United States Department of Education.

All data, study materials, analysis code, and preregistrations are publicly available at [https://researchbox.org/157&PEER\\_REVIEW\\_passcode=DMVAKF](https://researchbox.org/157&PEER_REVIEW_passcode=DMVAKF). This work has not been presented at a conference. Both co-authors contributed to all aspects of the research and approved the final version of the article. The authors declare no competing interests or conflicts.

Correspondence concerning this article should be addressed to Charles A. Dorison, Department of Management and Organizations, Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208, United States. Email: [charles.dorison@kellogg.northwestern.edu](mailto:charles.dorison@kellogg.northwestern.edu)

“replicate beyond any reasonable thresholds” (Ruggeri et al., 2020). Taken together, the weight of empirical evidence makes clear that loss–gain framing effects on risk preferences are among the most robust and influential phenomena in all of social and behavioral science.

Traditionally, such loss–gain decision frames are treated as irrelevant features of a decision that *should* be ignored when forming risk preferences—even though they are influential in practice (for dissenting opinions who argue that deviations from rational models can lead to more accurate judgments, see Gigerenzer & Goldstein, 1996; Gigerenzer & Gaissmaier, 2011; see also Sher & McKenzie, 2006). In their seminal work, Tversky and Kahneman (1981, p. 453) argued that framing effects violate the decision theory axioms of consistency and coherence. In his Nobel Prize address, Kahneman (2002) observed that “invariance is an essential aspect of rationality, which is violated in demonstration of framing effects.” Finally, in a more recent review, Bazerman and Moore (2009; p. 65) echoed a similar sentiment, concluding that “rational decision makers should be immune to the framing of choices.”

Taken together, prior research makes two points clear. First, on a descriptive level of analysis, framing effects are incredibly robust and influential. Second, on a prescriptive level of analysis, they should not be.

## Present Aims

In the present work, we qualify this strong prescriptive claim by documenting the reputational consequences of ignoring decision frames.<sup>1</sup> We begin from the foundational assumption that decision makers hold deep impression management concerns and care heavily about their reputations (Baumeister & Leary, 1995; Goffman, 1959; Leary et al., 2015; Lerner & Tetlock, 1999; Schlenker & Weigold, 1992; Tetlock, 2000, 2002). In many contexts (e.g., for an employee in the days before an important work promotion evaluation or for a public leader in the days before an election), decision makers may care more deeply about the immediate impressions they leave on others than the long-run expected value of their choices. Of note, this attention to reputation serves a social function in that one’s reputation plays a central role in determining others’ willingness to interact and cooperate with us (e.g., Jordan et al., 2016). Indeed, reputation may be especially critical for leaders, whose ability to successfully lead their teams and organizations depends on others’ perceptions of their competence, warmth, morality, and other related traits (e.g., Krels et al., 2017).

Drawing on this foundational assumption, we theorized that although ignoring decision frames is rational under the parsimonious assumptions of traditional choice models (Briggs, 2019; Von Neumann & Morgenstern, 1944), the unqualified claim that they should always be ignored is incomplete because it does not consider these critical reputational incentives. Specifically, we hypothesized that while third-party observers would penalize decision makers who make risk-averse (vs. risk-seeking) choices when choice outcomes are framed as losses, this result would reverse when outcomes are framed as gains. We tested this overarching hypothesis across five social perceptions (competence, warmth, confidence, morality, leadership effectiveness), three decision contexts (classic influenza problem, monetary gambles, COVID-19 policy response), two sample populations of United States adults

(Amazon Mechanical Turk, Harvard Digital Labs), and with real financial stakes.

Our results hold implications not only for loss–gain framing effects on risk preferences, but also for the broader interdisciplinary field of judgment and decision making. Traditionally, deviations from prescriptive models are considered suboptimal and in need of correction (but see also: Gigerenzer & Goldstein, 1996; Gigerenzer & Gaissmaier, 2011; Sher & McKenzie, 2006). Adding to other recent work in the behavioral sciences (e.g., Everett et al., 2018; Grossmann et al., 2020; Jordan et al., 2016; Tenney et al., 2019), the present research suggests that decision biases that appear suboptimal can be simultaneously reputationally rewarding. Reputational incentives may contribute to the persistence of such biases—given that they are reputationally reinforced—suggesting that interventions that target reputation may be an effective way for motivated individuals and organizations to reduce or eliminate these behaviors. Specifically, our results suggest a necessary focus on the social environment and cultural norms in addition to the cognition of the decision maker. We discuss these preliminary implications further in the General Discussion.

## Theoretical Background

Our theorizing takes a two-step approach. We hypothesized (a) that observers themselves fall victim to loss–gain effects on risk preferences and (b) that observers punish decision makers who make different choices than they themselves do. We review prior literature motivating these hypotheses below. Although both hypotheses draw directly from large prior literatures, they serve as the foundation for the novel contribution of the present article: that tendencies that have long been considered suboptimal (e.g., loss–gain framing effects on risk preferences) can serve a social function when considering critical reputational incentives.

### *Ignoring Decision Frames Increases Disagreement*

In line with a voluminous prior literature, we reasoned that third-party observers themselves typically fall victim to framing effects. For example, in the canonical influenza problem (described in detail in Experiment 1), Tversky and Kahneman (1981) found that whereas 72% of individuals made risk-averse choices when outcomes were framed as gains, just 22% of individuals made the same choice when outcomes were framed as losses, even though the options were logically equivalent.

It follows that decision makers who ignore frames will be more likely to make a choice that third-party observers disagree with than will individuals who attend to decision frames. Specifically, individuals who make *frame-consistent choices* (i.e., risk-averse choices in the gain frame or risk-seeking choices in the loss frame) will tend to disagree with third-party observers less often. Thus, a decision maker who ignores frames—and makes risk-seeking or

<sup>1</sup>The present article does not directly address the related descriptive claim of why decision makers fall victim to loss–gain framing effects on risk preferences in the first place. Prior research focuses on cognitive mechanisms to explain why framing effects occur. In the General Discussion, we consider whether reputational incentives may play an underappreciated role in reinforcing such behaviors.

risk-averse choices regardless of frame—will disagree with third-party observers more often than a decision maker who attends to frames.

### **Observers Punish Disagreement**

Multiple literatures both within and outside psychology suggest that disagreement will in turn drive reputational penalties. In contexts where people hold strong attitudes (e.g., the political domain), a large prior literature provides evidence for homophily: the tendency for partisans to prefer to affiliate with others who share their political beliefs (Dehghani et al., 2016; DiPrete et al., 2011; Halberstam & Knight, 2016; Huber & Malhotra, 2017; Kossinets & Watts, 2009; McPherson et al., 2001). In partisan conflict, observers not only perceive ingroup (vs. outgroup) members as high on both warmth and competence (Fiske, 2015), but also trust and cooperate more with them in economic games (Foddy et al., 2009; Rand et al., 2009; for review, see Minson & Dorison, 2021).

However, such effects can persist even in contexts where people do not hold strong, identity-relevant attitudes. As far back as 60 years ago, Byrne (1961) predicted and found not only that strangers are liked better when they are known to have similar (rather than dissimilar) attitudes to a focal participant, but also that they are perceived as more intelligent and moral. A decade later, Tajfel (1970) revealed that even artificial and insignificant differences (e.g., estimating the number of dots on a page) were sufficient to lead participants to give more money to members of their alleged group than to members of the other group.

Of note, such preferences for similar over dissimilar others have deep roots not only in psychology, but also in ancient philosophy. Centuries ago, in approximately 370 BCE in his dialogue *Phaedrus*, Plato wrote that “similarity begets friendship” (Plato, 1968, p. 837; see also Prinstein & Dodge, 2008). Indeed, the English proverb “birds of a feather flock together”—meaning that people with mutual attributes or preferences tend to affiliate—can be traced back to the 2nd century BCE.

Why might such effects hold in the context of disagreement over risk preferences? One possibility is that such preferences for similar over dissimilar others may be especially pronounced and robust due to halo effects (i.e., the influence of a global evaluation on evaluations of individual attributes of a person; Nisbett & Wilson, 1977). In a classic study on the halo effect, Nisbett and Wilson (1977) revealed that global evaluations of an individual (in this case, a college instructor’s warm and friendly vs. cold and distant manner) were powerful enough to alter evaluations of even relatively unambiguous stimuli (in this case, the college instructor’s appearance). In the present context, it could be the case that global evaluations of similarity (based on agreement vs. disagreement) are powerful enough to alter evaluations not only of choice-relevant attributes (e.g., competence), but also of relatively unrelated attributes.

A second possible explanation (although one that is less parsimonious) is rooted in naïve realism (Pronin et al., 2004; Robinson et al., 1995; Ross & Ward, 1995), sometimes called the “objectivity illusion” (Ross, 2018). Prior research on naïve realism theorizes that individuals consider their views and beliefs to be true representations of reality. To the extent that another individual sees things differently, they will be attributed negative traits such as being uninformed or biased (for review, see Pronin et al., 2004; Ross, 2018).

Whereas initial evidence for the naïve realism hypothesis was rooted in the domains of partisan conflict and negotiations (e.g., Robinson et al., 1995), researchers testing this hypothesis have also branched out to domains in which individuals do not hold strong, identity-relevant attitudes. For example, Minson et al. (2011) found that individuals gave less weight to a partner’s estimate than they did to their own in quantitative judgment tasks, presumably because they made negative inferences regarding the quality of the disagreeing partner’s advice (see also Liberman et al., 2012; Minson & Mueller, 2012).

Given that explanations based on halo effects and naïve realism make the same prediction regarding the relationship between disagreement and reputational penalties, the present experiments remain agnostic to their relative contribution (while noting that explanations based on halo effects are more parsimonious). However, we consider possibilities for future research to tease apart these explanations in the General Discussion.

### **Summary**

Taken together, a large and long-standing body of work suggests that (1) decision makers who do not fall victim to framing effects will disagree with observers more often and (2) disagreement (vs. agreement) leads to negative reputational and social outcomes. Building off this diverse body of theorizing, we predicted that in the domain of risk preferences, disagreement over the “right” decision would underpin (i.e., statistically mediate) negative inferences for decision makers who did not fall victim to loss–gain framing effects on risk preferences.

### **Research Overview**

We conducted three preregistered experiments (collective  $N = 1,954$ ) to test the hypothesis that while individuals who make risk-averse (vs. risk-seeking) choices will be socially and financially penalized by third-party observers when decisions are framed as losses, this pattern will reverse when decisions are framed as gains. Critically, our main claim is the assertion that decision makers purposefully fall victim to loss–gain framing effects on risk preferences to make a positive impression on observers (i.e., focused on reputational causes). Rather, our main claim is focused on reputational consequences. Specifically, we hypothesize that when decision makers follow the prescriptive advice to ignore decision frames, they incur social and economic penalties. Thus, in our experiments, we focus on judgments made by third-party observers rather than choices by decision makers themselves, although we discuss preliminary extensions to decision makers in the General Discussion.

### **Open Science Statement**

We report how we determined our sample size, all exclusions, all manipulations, and all measures in all experiments (Simmons et al., 2012). All experiments were preregistered on [aspredicted.org](https://aspredicted.org) and all preregistrations, materials, data, and code are available on [researchbox.org](https://researchbox.org).<sup>2</sup>

<sup>2</sup> Researchbox: [https://researchbox.org/157&PEER\\_REVIEW\\_passcode=DMVAKF](https://researchbox.org/157&PEER_REVIEW_passcode=DMVAKF), Experiment 1: <https://aspredicted.org/blind.php?x=dd7xs4>, Experiment 2: <https://aspredicted.org/blind.php?x=6kb6vb>, Experiment 3: <https://aspredicted.org/blind.php?x=22dq47>

## Experiment 1

Experiment 1 tested whether ignoring decision frames carries reputational costs in the canonical influenza problem (Tversky & Kahneman, 1981; Problems 1 & 2). Experiment 1 was conducted in November 2019, before the COVID-19 pandemic. Participants in this experiment played the role of observer, in which they reported their impressions of a target decision maker after observing their choice.

### Method

Participants (i.e., observers) were randomly assigned to one of four between-subjects experimental conditions in a 2 (Frame: loss, gain)  $\times$  2 (Target choice: risk-averse, risk-seeking) design, as described below. All experiments received Institutional Review Board approval and participants gave informed consent in all experiments.

### Participants

We aimed to recruit 440 individuals living in the United States from Amazon's Mechanical Turk through CloudResearch. We advertised the experiment as a "survey about judgment and decision making." Sample size yielded 80% power to detect moderate effect sizes (i.e., Cohen's  $d$  between .30 – .50) for simple effects. Owing to chance factors, we ended up with 448 participants (199 women; mean age = 36.90, age range = 19–70). In accordance with our preregistration, we excluded data from 67 participants who missed any of the three comprehension check questions asking about the details of the scenario. This left us with a total of 381 participants. In this and all experiments, our results remain consistent regardless of whether we implement the exclusions (see online materials on Researchbox.org).

### Procedure

Participants read a third-person adaptation of the classic influenza problem initially conceived by Tversky and Kahneman (1981; see also Kahneman & Tversky, 1979). Participants were told to imagine that "Casey is in charge of the United States Centers for Disease Control and Prevention. The United States is preparing for the outbreak of a new strain of the flu, which is expected to kill 600 people in this country." All participants were then told that Casey faced a decision between two alternative programs, with the framing of the outcomes associated with each program's adoption varying by condition.

As in past work on the framing effect, participants were then randomly assigned to either the gain condition or the loss condition. In the gain condition, participants were told, "If Program A is adopted, 200 people will be saved. If Program B is adopted, there is a one third probability that all 600 people will be saved and a two thirds probability that no one will be saved." In the loss condition, participants were told, "If Program A is adopted, 400 people will die. If Program B is adopted, there is a one third probability that nobody will die and a two thirds probability that 600 people will die." Notably, although the programs are equivalent from a prescriptive decision theory perspective (i.e., 200 people saved is equivalent to 400 people dying when the population is 600 people), a mature literature provides robust evidence that individuals typically prefer the certain option in the gain frame but prefer the

risky option in the loss frame (for review, see Ruggeri et al., 2020). After reading the scenario, participants answered three comprehension checks. Then, participants indicated what they themselves would choose to do if they were in Casey's position.

After answering the comprehension questions and indicating their own choice, participants were told what Casey had chosen to do in the scenario. Participants were randomly assigned to either the risk-averse condition (in which Casey chose Program A) or the risk-seeking condition (in which Casey chose Program B).

### Social Perceptions

Whereas prior work on the framing effect typically uses participant choices as the primary dependent variable, our dependent variables were social perceptions of Casey (i.e., the target decision maker). After seeing Casey's choice, participants rated Casey on three confirmatory dependent variables, which were chosen to assess a mix of distinct facets of reputation relevant to leadership: perceived competence, warmth, and confidence (Fast et al., 2012; Fiske et al., 2002). Order of presentation for the three variables was randomized. To measure competence and warmth (Fiske et al., 2002), participants rated how well nine scale items (five for competence, four for warmth; e.g., "Intelligent"; "Tolerant") fit the decision maker on a 5-point scale (1 = *not at all*, 5 = *extremely*). To measure confidence, participants rated their level of agreement with four statements describing Casey (e.g., "is very sure about what he knows") on a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*). We adapted the confidence in thoughts and feelings scale from prior work (Fast et al., 2012; see also John et al., 2019). Because the competence, warmth, and confidence scales all achieved sufficient levels of reliability (Cronbach's  $\alpha$  = .79, .90, and .92, respectively), we averaged scale items to compute scores for each given scale.

After the three confirmatory dependent variables, we also measured two exploratory social perceptions relevant to leadership: perceived morality and general leadership effectiveness (e.g., Kreps, Laurin, & Merritt, 2017). We included the first three variables (competence, warmth, confidence) as outcome measures in confirmatory hypotheses because we measured each construct of interest with a validated scale from the prior literature. We included the latter two (leadership and morality) as outcomes in exploratory hypotheses because we measured both constructs with single-item Likert measures that were not validated by prior research (to our knowledge). Both were single 7-point Likert items (i.e., "To what extent, if at all, do you think Casey is moral?" and "To what extent, if at all, do you think Casey is an effective leader?") and were answered in a randomized order after the completion of the three confirmatory dependent variables.

### Demographics

After completing all other questions, participants reported demographic information (age, gender, education) and whether they had previously been taught the principle of the framing effect.

## Results

### Moderation Analyses

We first tested our key hypothesis: whether the reputationally optimal choice for the target decision maker depended on the

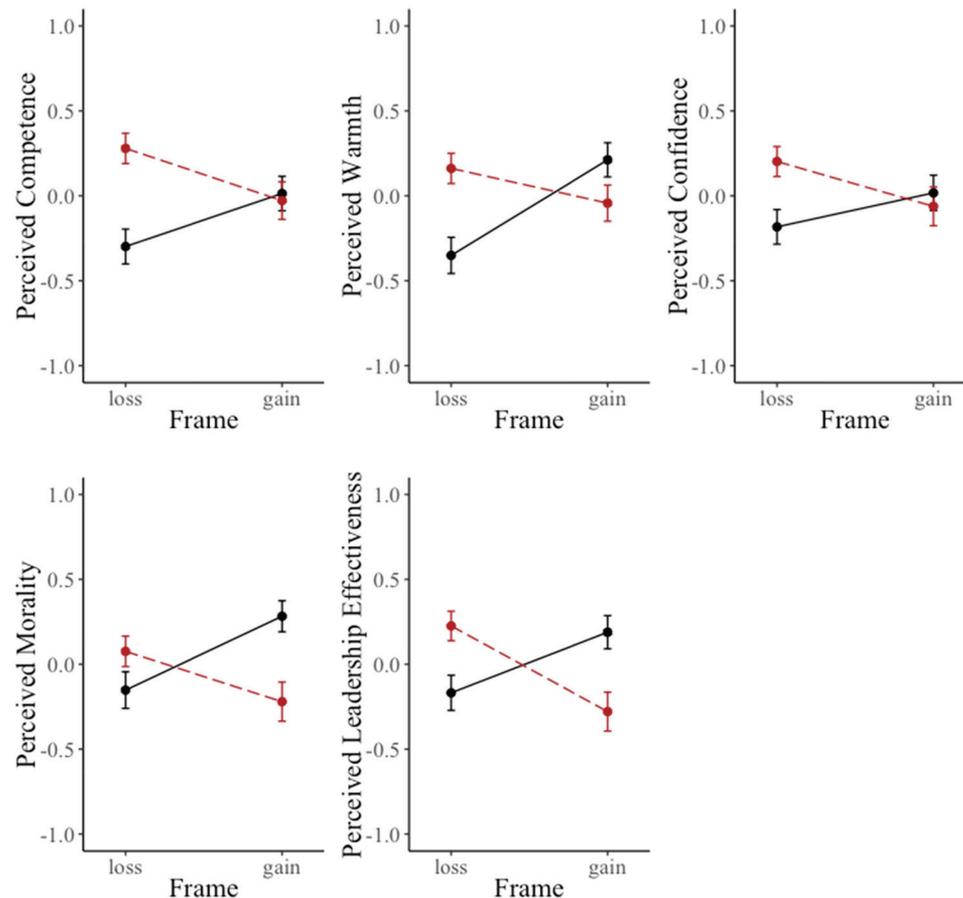
frame in which the choice was made. On the one hand, it could be the case that observers would always perceive the risk-averse target or the risk-seeking target more positively, regardless of the frame in which the options were presented. On the other hand, in line with our theorizing, it could be the case that social perceptions follow directly from third-party observers' own preferences, which are themselves influenced by decision frames.

To test these competing hypotheses, we fit three separate regressions in which we regressed our confirmatory social perception variables on frame (loss, gain), target choice (risk-averse, risk-seeking), and their interaction. Our key hypothesis test was the significance of the interaction term. Because our dependent variables were measured on different scales (some 5-point Likert scales and some 7-point Likert scales), we standardized each dependent variable to have a mean of zero and standard deviation of one. We standardized the dependent variable for all moderation analyses in

Experiments 1–2 to facilitate comparison across results given their similar designs (we also report standardized effect sizes in Experiment 3, detailed later).

As predicted, we found a significant interaction between frame and target choice for all three confirmatory dependent variables: perceptions of competence ( $b = -.62$ , 95% CI  $[-1.02, -.22]$ ,  $t = -3.08$ ,  $p = .002$ ), warmth ( $b = -.77$ , 95% CI  $[-1.16, -.37]$ ,  $t = -3.82$ ,  $p < .001$ ), and confidence ( $b = -.46$ , 95% CI  $[-.87, -.06]$ ,  $t = -2.28$ ,  $p = .023$ ). We also found a significant interaction for our two exploratory dependent variables related to leadership: perceptions of morality ( $b = -.73$ , 95% CI  $[-1.13, -.33]$ ,  $t = -3.62$ ,  $p < .001$ ) and leadership effectiveness ( $b = -.86$ , 95% CI  $[-1.26, -.47]$ ,  $t = -4.23$ ,  $p < .001$ ). Of note, and as visualized in Figure 1, we found minimal evidence of asymmetry in our effects across the gain versus loss frames: in all cases, the risk-averse target was directionally preferred in the gain frame while the risk-

**Figure 1**  
*Interaction Between Frame (Loss, Gain) and Partner Choice (Risk-Seeking, Risk-Averse) on Five Different Social Perceptions in Experiment 1 (N = 381)*



*Note.* Solid black bars are evaluations of a risk-averse target, and dashed red bars are evaluations of a risk-seeking target. Error bars represent one standard error. The dependent variable was standardized in all cases. In all cases, social perceptions depended both on the choice and the frame in which the options were presented. Whereas risk-seeking choices were socially rewarded when options were framed as losses, risk-averse choices were socially rewarded when options were framed as gains. See the online article for the color version of this figure.

seeking target was directionally preferred in the loss frame. Further results are presented in Table 1.

Taken together, this set of results demonstrates that the reputationally-optimal choice depended on the frame in which the choice was made, providing initial evidence that decision makers who hold impression management concerns should not always be immune to framing effects.

To put these results in perspective, we ran a simulation in which we randomly drew 10,000 pairs of participants, one who evaluated a frame-consistent decision maker (i.e., risk-seeking in the loss frame or risk-averse in the gain frame) and one who evaluated a frame-inconsistent decision maker (i.e., risk-averse in the loss frame or risk-seeking in the gain frame). For each pair, we then assessed how often the frame-consistent decision maker was evaluated more positively than the frame-inconsistent decision maker (McGraw & Wong, 1992). Participants who evaluated a frame-consistent decision maker gave higher ratings of competence than their randomly selected match who evaluated a frame-inconsistent decision maker 55% of the time, and the reverse just 36% of the time (the remaining 9% of pairs indicated equal perceptions of competence). A similar pattern arose for perceived warmth (57% vs. 32% vs. 11%), confidence (52% vs. 39% vs. 9%), morality (47% vs. 30% vs. 23%), and leadership effectiveness (51% vs. 27% vs. 22%). These simulations reinforced the consistent reputational reward associated with making frame-consistent choices, despite the logical equivalence of the choices across frames.

### Mediation Analyses

A key remaining question was whether agreement with the choice of the decision maker underpinned the interaction we observed regarding social perceptions. To answer this question, we fit five separate moderated mediation models using the Lavaan package in R (Rosseel, 2012). The models were equivalent to Hayes Process model 7 (Hayes, 2017). A sample model is depicted in Figure 2. Summary results for all five models, including 95% confidence intervals for the index of moderated mediation and percentage of the total effect mediated in each model, are presented in Table 2.

All five models tested the same overarching moderated mediation hypothesis. In all five models, the independent variable was the choice made by the target decision maker (1 = risk-seeking, 0 = risk-averse). The mediating variable in all five models was agreement between the target's choice and the observer's own preference (1 = target's choice was the same as the observer's, 0 = target's choice was different than the observer's) and the moderating

variable on the "a" path was the frame (1 = gain frame, 0 = loss frame). Finally, our dependent variables were (standardized) perceptions of competence, warmth, confidence, morality, and leadership effectiveness in Models 1–5, respectively.

In all five models, we found evidence consistent with the hypothesized moderated mediation (indices of moderated mediation =  $-.27$ ,  $-.29$ ,  $-.21$ ,  $-.34$ , and  $-.42$ , respectively, all  $ps < .003$ ). First, we found a significant interaction between the target decision maker's choice and frame on agreement ( $b = -.52$ , 95% CI  $[-.71, -.33]$ ,  $t = -5.34$ ,  $p < .001$ , identical in all five models). This interaction occurred because observers themselves fell victim to the framing effect, increasing rates of agreement with frame-consistent target decision makers. Specifically, observers in the loss frame preferred the risky choice 57% of the time, while participants in the gain frame preferred the risky choice only 30% of the time. Second, agreement positively predicted social perceptions in all five models ( $bs = .51$ ,  $.56$ ,  $.41$ ,  $.66$ , and  $.81$ , respectively, all  $ps < .001$ ).

Taken together, across the five moderated mediation models, the results provided evidence consistent with the hypothesis that agreement with the target decision maker's choice itself underpinned social perceptions. These results were consistent with the theory that observers socially derogate decision makers with diverging preferences.

### Individual Difference Analyses

Finally, we conducted exploratory analyses to examine whether any individual difference variables attenuated (or amplified) our moderation results. Specifically, we tested whether any of our four individual difference variables (age, gender, education, self-reported knowledge of framing effects) moderated any of our five two-way interactions (competence, warmth, confidence, leadership, morality).

We found minimal evidence for moderation based on individual difference variables. Of the twenty three-way interactions tested, only one was statistically significant: higher levels of education significantly attenuated the two-way interaction for perceived confidence ( $b = -.37$ , 95% CI  $[-.71, -.03]$ ,  $t = -2.13$ ,  $p = .034$ ). This result did not hold after accounting for multiple hypothesis testing and was likely due to chance factors given the large number of three-way interactions tested, but it could perhaps provide an interesting avenue for future research.

As a whole, the exploratory individual difference analyses suggested that our results were not driven by any particular subpopulation and were broadly generalizable among the subgroups we considered in the sample.

**Table 1**

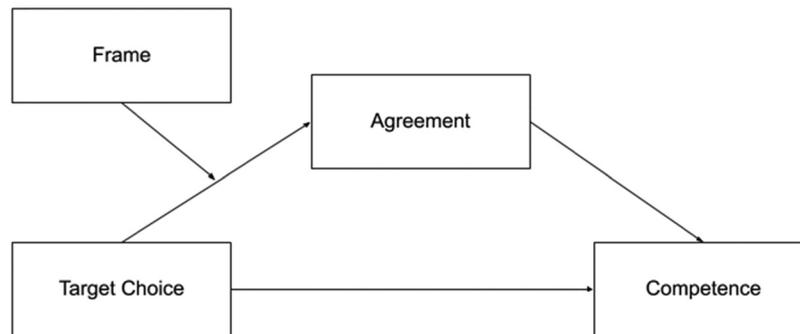
Regression Results for Five Interaction Models in Experiment 1 ( $N = 381$ )

Measure	Dependent variable				
	Competence (1)	Warmth (2)	Confidence (3)	Leadership (4)	Morality (5)
Risky choice	0.578*** (0.140)	0.512*** (0.140)	0.385** (0.142)	0.394** (0.140)	0.228 (0.141)
Gain frame	0.312* (0.144)	0.563*** (0.144)	0.200 (0.146)	0.357* (0.144)	0.435** (0.144)
Choice × Frame	-0.620** (0.202)	-0.767*** (0.201)	-0.464* (0.204)	-0.861*** (0.201)	-0.731*** (0.202)
Constant	-0.299** (0.102)	-0.351*** (0.102)	-0.183 (0.103)	-0.168 (0.102)	-0.152 (0.102)
Observations	381	381	381	381	381
R <sup>2</sup>	0.043	0.049	0.020	0.048	0.038

Note. The dependent variable was standardized in all cases. In all five models, there was a significant interaction between frame (loss, gain) and partner choice (risk-seeking, risk-averse).

\*  $p < .050$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

**Figure 2**  
One of Five Moderated Mediation Models Fit in Experiment 1 ( $N = 381$ )



*Note.* In all five models, the independent variable was target choice, the mediating variable was agreement, and the moderating variable was frame. The dependent variables were standardized competence, warmth, confidence, morality, and leadership effectiveness in Models 1–5, respectively. We fit the identical model in Experiment 2 ( $N = 369$ ) with dictator game giving, competence, and warmth as the dependent variables in Models 1–3, respectively.

## Discussion

Experiment 1 provided evidence that ignoring decision frames carries reputational costs in the canonical influenza problem because observers themselves fell victim to framing effects and punished decision makers who disagreed. We found minimal evidence for further moderation by individual difference variables.

## Experiment 2

Experiment 2 tested the same overarching hypotheses with another canonical scenario and with real financial stakes. Experiment 2 extended Experiment 1 in at least three important ways. First, the scenario now pertained to monetary gambles instead of public policy choices, allowing us to assess whether social penalties for ignoring decision frames were limited to life-or-death decisions or whether the pattern of results from Experiment 1 generalized beyond the specific scenario considered therein. Second, we now implemented a dictator game with real financial stakes, in addition to social judgments. Finally, whereas Experiment 1 (and Experiment 3, below) examines the effect of describing an identical gamble differently relative to a reference point (i.e., a framing effect),

Experiment 2 examines the effect of gambles differing in the sign of the outcome itself (i.e., a reflection effect; see Fagley, 1993).

We predicted that—even with a new scenario, real money at stake, and in the context of a reflection effect—the effect of target choice on observer judgments would depend not only on the choice but also on the frame in which the choice was made. Finally, we again theorized that agreement with the choice itself would underpin such reward.

## Method

As in Experiment 1, participants were randomly assigned to one of four between-subjects experimental conditions in a 2 (Frame: Loss, Gain)  $\times$  2 (Partner preference: risk-averse, risk-seeking) design. Participants again played the role of a third-party observer. The scenario, described in greater detail below, was an adaptation of Problems 3 and 4 from Tversky and Kahneman (1981).

## Participants

Experiment 2 was conducted in January 2020. We aimed to recruit 440 individuals living in the United States from Amazon’s Mechanical Turk through CloudResearch. We advertised the experiment as a “survey about judgment and decision making.”

**Table 2**  
Results of Moderated Mediation Analyses in Experiment 1 ( $N = 381$ )

Dependent variable	Moderated mediation		
	Index of moderated mediation	95% confidence interval	Percent of total effect mediated
Competence	–0.27	[–0.43, –0.13]	25%
Warmth	–0.29	[–0.47, –0.15]	22%
Confidence	–0.21	[–0.36, –0.09]	27%
Morality	–0.34	[–0.53, –0.19]	28%
Leadership	–0.42	[–0.63, –0.24]	29%

*Note.* In all five models, the independent variable was target choice, the mediating variable was agreement, and the moderating variable was frame. The dependent variables were standardized competence, warmth, confidence, morality, and leadership effectiveness in Models 1–5, respectively.

Sample size again yielded 80% power to detect moderate effect sizes (i.e., Cohen's  $d$  between .30–.50) for simple effects. Owing to chance factors, we ended up with 444 participants (161 women; mean age = 35.53, age range = 19–72). In accordance with our preregistration, we excluded data from 75 participants who missed any of the two comprehension check questions asking about the details of the scenario. This left us with a total of 369 participants.

### Procedure

Upon entering the experiment, we told participants that in addition to the base payment for the experiment, they would receive a \$.20 bonus to allocate between themselves and another MTurk worker, known as the Receiver. Participants were informed that they could decide how much, if any, of the available bonus funds they wanted to share with the Receiver based on how well they thought the Receiver performed in a scenario later in the experiment; anything they did not award, they could keep for themselves. In reality, the Receivers were fictitious and their actions were randomly assigned; however, participants kept any money that they did not send to the Receiver.

Next, participants read a third-party adaptation of “Problem 3” or “Problem 4” from Tversky and Kahneman (1981). While the scenario in Experiment 1 presented decision makers with a decision about saving lives, this experiment focused on monetary gambles. Participants in the gain frame were given the following options: (Option A) a sure gain of \$240 versus (Option B) a 25% chance to gain \$1,000 and a 75% chance to gain nothing. Participants in the loss frame were given the following options: (Option A) a sure loss of \$750 versus (Option B) 75% chance to lose \$1,000 and a 25% chance to lose nothing. After reading the scenario, participants answered two comprehension checks regarding the details of the scenario and then reported their own preference for what they themselves would choose.

On the following page, participants in the risk-averse condition were told that their partner chose Option A (i.e., the sure thing), whereas participants in the risk-seeking condition were told that their partner chose Option B (i.e., the gamble).

### Dictator Game

After being told their partner's choice, participants played a dictator game with their partner. Participants were told,

It is up to you to decide how much, if any, of the available bonus funds you want to share with the Receiver based on how well they made their decision. Whatever you do not award to the Receiver, you will keep for yourself.

Participants were asked, “How many cents will you award to the Receiver based on their decision?” before dragging a slider between 0 and 20 cents to choose how much of the bonus funds to send to the receiver and how much to keep for themselves. The slider defaulted to 10 cents (i.e., 50%), such that any amount below 10 cents represented a punishment and any amount above 10 cents represented a reward.

### Social Perceptions

Following the dictator game, we collected two further social perceptions: competence and warmth (alphas = .74 and .86, respectively).

We used the same measures as in Experiment 1. We report a summary of the results for these secondary analyses in the results section below and in full in the online materials on Researchbox.org.

### Demographics

After completing all other variables, participants reported demographic information (age, gender, education) and whether they had previously been taught the principle of the framing effect.

## Results

### Moderation Analyses

As in Experiment 1, we first tested our key confirmatory hypothesis: whether the reputationally optimal choice depended on the frame in which the choice was made. The primary dependent variable in this experiment was the percentage of an endowment (\$.20) shared with the Receiver (i.e., their allegedly real partner) in a dictator game. As in Experiment 1, we standardized all dependent variables in all moderation analyses to facilitate comparison across results.

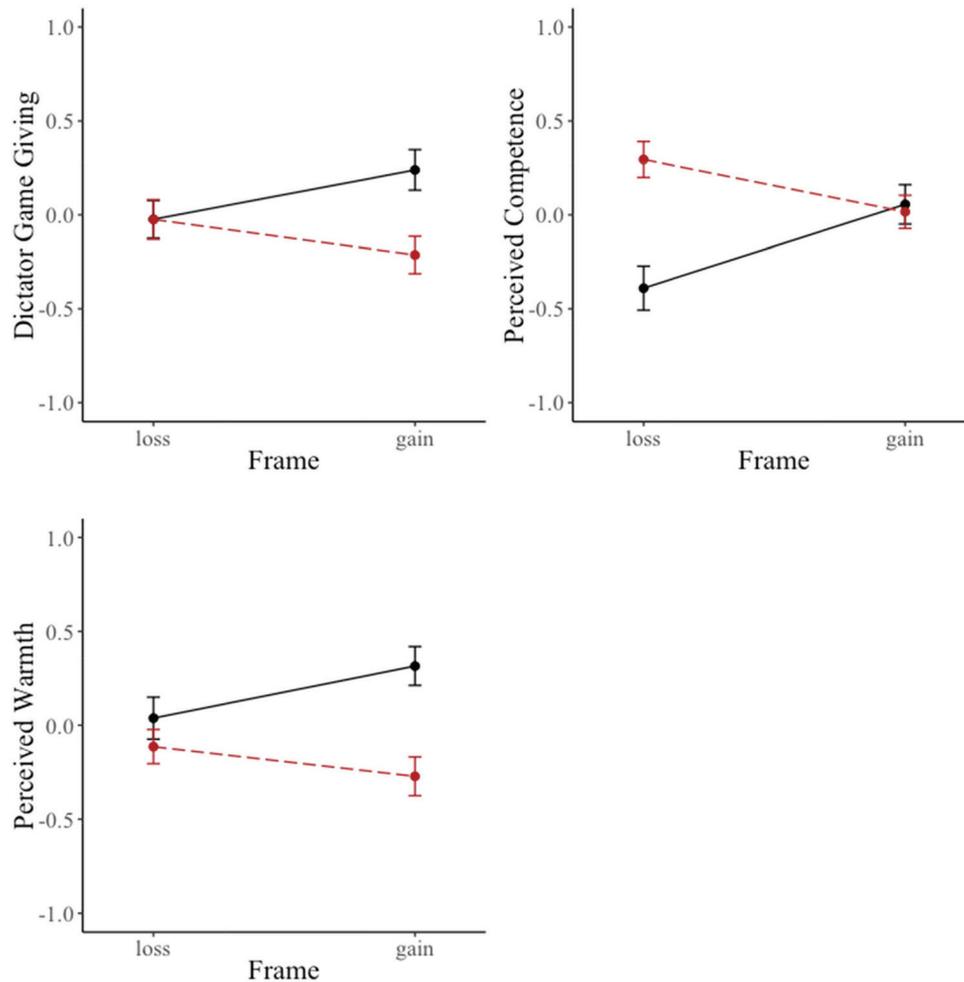
Following the same empirical strategy as Experiment 1, we tested this hypothesis by regressing dictator game behavior on frame (loss, gain), partner choice (risk-averse, risk-seeking), and their interaction. As predicted, we found a significant interaction between frame and choice on the amount of money sent to a partner ( $b = -.45$ , 95% CI  $[-.86, -.04]$ ,  $t = -2.18$ ,  $p = .030$ ). For ease of interpretation, we also report the impact of frame and partner choice on the nonstandardized percentage of endowment sent to a partner ( $b = -14.8%$ , 95% CI  $[-28.1%, -1.4%]$ ) and the number of cents (out of a maximum of 20) sent to a partner ( $b = -2.95$ , 95% CI  $[-5.63, -.28]$ ).

In a secondary analysis, we also replicated our results from Experiment 1: We found a significant interaction between frame and partner choice on both perceptions of competence ( $b = -.72$ , 95% CI  $[-1.13, -.32]$ ,  $t = -3.55$ ,  $p < .001$ ) and warmth ( $b = -.44$ , 95% CI  $[-.84, -.03]$ ,  $t = -2.13$ ,  $p = .034$ ). Results are visualized in Figure 3 and presented in Table 3.

To put these results in perspective, we conducted an exploratory analysis to assess how much money observers shared with partners who made a frame-inconsistent choice (i.e., risk-averse in the loss frame or risk-seeking in the gain frame) compared with partners who made a frame-consistent choice (i.e., risk-seeking in the loss frame or risk-averse in the gain frame). Whereas observers shared 27% of their endowment with participants who made frame-inconsistent choices, they shared 35% of their endowment with those who made frame-consistent choices. This difference was statistically significant ( $b = 7.64$ , 95% CI  $[.94, 14.34]$ ,  $t = 2.24$ ,  $p = .026$ ) and was indicative of a relative increase of almost 30% (an approximately 8% absolute increase on a 27% baseline). Results are presented in Figure 4.

Contrary to our expectations, in this experiment (but not in Experiment 1 above or in Experiment 3 below), an exploratory analysis of the simple effects revealed that the focal interaction for dictator game giving was driven primarily by behavior in the gain frame. Further exploratory analyses, which we report in full in the online materials on Researchbox.org, provide evidence that third-party observers financially differentiated between partners in the gain frame (but not the loss frame) because (a) they differentiated between partners on perceptions of warmth in the gain frame (but

**Figure 3**  
Interaction Between Frame (Loss, Gain) and Partner Choice (Risk-Seeking, Risk-Averse) on Three Different Dependent Variables in Experiment 2 ( $N = 369$ )



*Note.* Solid black bars are evaluations of a risk-averse target, and dashed red bars are evaluations of a risk-seeking target. Error bars represent one standard error. The dependent variable was standardized in all cases. In all cases, social perceptions depended both on the choice and the frame in which the options were presented. See the online article for the color version of this figure.

not the loss frame) and (b) perceptions of warmth were closely tied to dictator game giving in this experiment.

### Mediation Analyses

We next examined whether agreement with the partner's choice underpinned financial reward in the dictator game. Our analytic strategy followed the exact strategy from Experiment 1, except our dependent variable was the (standardized) amount of money sent to the partner.

We again found evidence consistent with the hypothesized moderated mediation (index of moderated mediation =  $-.23$ , 95% CI  $[-.44, -.02]$ ,  $t = -2.09$ ,  $p = .037$ ). First, we found a significant interaction between target choice and frame on agreement ( $b = -1.00$ , 95% CI  $[-1.16, -.83]$ ,  $t = -11.79$ ,  $p < .001$ ). As in Experiment 1, this occurred because observers themselves fell victim to the framing

effect: observers in the loss frame preferred the risk-seeking choice 58% of the time, whereas observers in the gain frame preferred the risk-seeking choice only 8% of the time. Second, agreement positively predicted dictator game giving ( $b = .23$ , 95% CI  $[.02, .43]$ ,  $t = 2.17$ ,  $p = .030$ ). Further, we again found consistent evidence for moderated mediation in this experiment using the same model but with competence and warmth as the dependent variables (indices of moderated mediation =  $-.72$  and  $-.37$ , respectively, both  $ps < .002$ ). Summary results for all three models, including 95% confidence intervals for the index of moderated mediation and percentage of the total effect mediated in each model, are presented in Table 4.

Taken together, the family of evidence from five moderated mediation models in Experiment 1 and three models in Experiment 2 provided evidence consistent with the hypothesis that agreement with a target decision maker's choice underpinned social and financial reward from third-party observers.

**Table 3**  
*Regression Results for Three Interaction Models in Experiment 2 (N = 369)*

Measure	Dependent variable		
	Dictator game giving (1)	Competence (2)	Warmth (3)
Risky choice	0.263 (0.146)	0.446** (0.144)	0.278 (0.144)
Gain frame	-0.001 (0.149)	0.685*** (0.147)	-0.151 (0.148)
Choice × Frame	-0.452* (0.208)	-0.725*** (0.204)	-0.436* (0.205)
Constant	-0.024 (0.107)	-0.390*** (0.105)	0.038 (0.106)
Observations	364	365	365
R <sup>2</sup>	0.027	0.058	0.049

*Note.* The dependent variable was standardized in all cases. In all three models, there was a significant interaction between frame (loss, gain) and partner choice (risk-seeking, risk-averse).

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

### Individual Difference Analysis

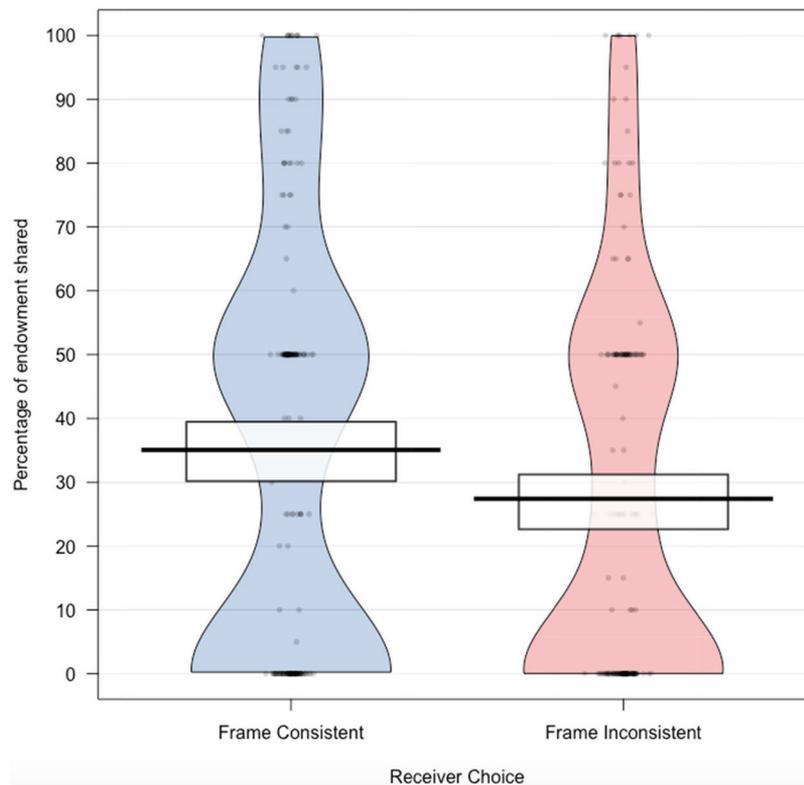
Finally, we conducted exploratory analyses of whether any of our four individual difference variables (age, gender, education, self-reported knowledge of framing effects) attenuated (or amplified) any of our three two-way interactions (dictator game giving, perceived competence, perceived warmth). Of these twelve three-way interactions tested, none reached statistical significance. As in Experiment 1, the exploratory individual difference analyses suggested that our results

were not driven by any particular subpopulation and were broadly generalizable among the subgroups we considered in the sample.

### Discussion

Experiment 2 replicated Experiment 1 with a new scenario and with real financial stakes. We again found minimal evidence for moderation by individual difference variables.

**Figure 4**  
*Observers Shared More of Their Endowment With Receivers Who Made Frame-Consistent Choices (Risk-Seeking in the Loss Frame or Risk-Averse in the Gain Frame) Compared With Receivers Who Made Frame-Inconsistent Choices (Risk-Averse in the Loss Frame or Risk-Seeking in the Gain Frame; Experiment 2, N = 369)*



*Note.* Violin plots display the distributions, and horizontal bars indicate the means. Rectangles show 95% confidence intervals. See the online article for the color version of this figure.

**Table 4**  
Results of Moderated Mediation Analyses in Experiment 2 ( $N = 369$ )

Dependent variable	Moderated mediation		
	Index of moderated mediation	95% confidence interval	Percent of total effect mediated
Dictator game giving	-0.23	[-0.44, -0.02]	51%
Competence	-0.72	[-0.98, -0.48]	100%
Warmth	-0.37	[-0.61, -0.15]	84%

*Note.* In all three models, the independent variable was target choice, the mediating variable was agreement, and the moderating variable was frame. The dependent variables were standardized dictator game giving, competence, and warmth in Models 1–3, respectively.

### Experiment 3

Experiment 3 addressed four further potential theoretical and methodological boundary conditions of Experiments 1 and 2. First, we tested the same overarching hypothesis in the context of COVID-19 rather than in a canonical framing effect scenario. While using canonical framing scenarios in Experiments 1 and 2 allowed us to most closely tie our hypotheses to prior literature, it could be the case that the results do not generalize to novel decision contexts such as an unfolding public health crisis. Second, participants now observed two targets making diverging choices rather than just a single target's choice. Prior research demonstrates that preferences can sometimes reverse in joint versus separate evaluations (Hsee et al., 1999). Third, whereas in prior experiments participants first indicated their own preference before observing the focal judgment or decision and evaluating target decision makers, in this experiment participants evaluated the decision makers without indicating their own preference. Finally, participants in both Experiments 1 and 2 were recruited via Mturk samples. It could be that prior results do not generalize outside this sample population. The present experiment recruited volunteer participants from the Harvard Digital Lab for the Social Sciences (DLABSS) to address this concern. Information about DLABSS can be found here: <http://dlabss.harvard.edu>.

### Method

Participants were randomly assigned to one of two between-subjects experimental conditions (Frame: Loss, Gain). Participants again played the role of a third-party observer. The scenario was an adaptation of Problems 1 and 2 from Tversky and Kahneman (1981) for the unfolding COVID-19 pandemic, described in greater detail below.

### Participants

Experiment 3 took place in July and August 2020. We recruited individuals living in the United States who had volunteered to take surveys with DLABSS. We advertised the experiment as a “survey about policy preferences during the COVID-19 pandemic.”

We preregistered to recruit 1,000 participants or for 10 weeks, whichever came first (checking recruitment in weekly increments). After 4 weeks, 1,105 participants had taken our survey and we thus terminated data collection (369 women; mean age = 54.59, age range = 19–92). In accordance with our preregistration, we included data from only the 906 participants who correctly answered all three attention checks. This yielded a sample size large enough to detect small effect sizes (i.e., Cohen's  $d$  of approximately .20) with over 80% power.

### Procedure

Participants read a version of the classic influenza problem used in Experiment 1, adapted to the context of COVID-19 policy. In both conditions, participants read the following: “Imagine that you are a citizen of a midsized U.S. state and are deciding between two candidates for governor of your state. The two candidates are proposing different plans to address the current COVID-19 pandemic. Assume that if neither candidate's plan is adopted, the COVID-19 pandemic is expected to kill 6,000 more people in your state.” We adapted the scenario to mirror consequential judgments that lay individuals regularly make outside of our specific experimental context (e.g., in the context of the November 2020 general election in the United States).

In the gain condition, participants then read:

If Candidate Smith is elected, he will implement a policy in which 2,000 people will be saved for sure. If Candidate Jones is elected, he will implement a policy in which there is a one third chance that 6,000 people will be saved and a two thirds chance that no one will be saved.

In the loss condition, participants instead read:

If Candidate Smith is elected, he will implement a policy in which 4,000 people will die for sure. If Candidate Jones is elected, he will implement a policy in which there is a one third chance that no one will die and a two thirds chance that 6,000 people will die.

Both (a) the order of the risk-averse and risk-seeking candidates and (b) the names of the candidates were counterbalanced.

### Social Perceptions

We collected five items measuring the perceived leadership ability of the candidates for governor. Participants indicated (a) which candidate they thought was a better leader; (b) which candidate they would vote for; (c) which candidate they thought would be a more effective governor; (d) which candidate they thought had better policies; and (e) which candidate they thought was more competent. All measures were answered on a 5-point scale ranging from  $-2$  (*Definitely Candidate Smith*) to  $+2$  (*Definitely Candidate Jones*), where a score of 0 indicated *no preference*.<sup>3</sup> Scores were recoded such that positive scores always indicated a preference for the risk-seeking candidate, whereas negative scores always indicated a preference for the risk-averse candidate. The 5-item measure achieved a high level of internal reliability (Cronbach's  $\alpha = .96$ ). We thus averaged the five items to create an index of perceived leadership ability.

<sup>3</sup> In our preregistration, we had a typo that indicated that these questions were answered on a 7-point scale.

### Individual Difference Variables

We collected a variety of individual difference variables for exploratory analyses. We highlight three measures of particular theoretical relevance here. First, we asked participants six questions measuring the tendency to engage in cognitive reflection (Frederick, 2005; Thomson & Oppenheimer, 2016). Second, we asked participants whether they had heard of the framing effect. Finally, we asked participants two questions about voting behavior: whether they were currently registered to vote in the United States and whether they planned to vote in the United States 2020 presidential election.

### Demographics

DLABSS collects demographic information from participants in a master survey that is administered the first time a volunteer participates in a study. DLABSS staff anonymously linked demographic data from the master survey to responses from Experiment 3. Information in the DLABSS master survey included gender, age, race, ethnicity, country of birth, zip code, religion, political affiliation, political orientation, language spoken at home, marital status, home ownership, occupation, annual income, and educational attainment. We did not analyze any of these variables in the present investigation, although they could be examined in future research.

## Results

### Main Effect Analyses

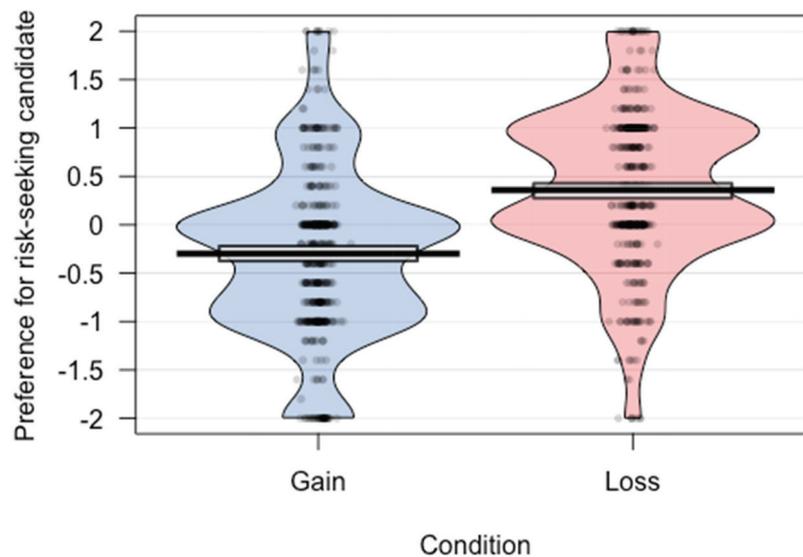
In this experiment, we did not standardize the leadership index because a raw score of 0 had a meaningful interpretation of “no preference.” Reported regression coefficients thus represent scale

points. However, we report standardized effect sizes (in the form of Cohen’s  $d$ ) for all regression analyses.

We first tested our key confirmatory hypothesis: that leaders who espoused risk-seeking policies (as compared with risk-averse policies) would be relatively more preferred when policies were framed as losses compared with when policies were framed as gains. As depicted in Figure 5, this was the case: Participants in the loss condition preferred the risk-seeking candidate to a greater extent than did participants in the gain condition ( $b = -.66$ , 95% CI  $[-.77, -.55]$ ,  $t = -11.66$ ,  $p < .001$ , Cohen’s  $d = .78$ ). Notably, we observed a full preference reversal: while participants significantly preferred the risk-seeking gubernatorial candidate in the loss frame ( $b = .36$ , 95% CI  $[.28, .44]$ ,  $t = 9.25$ ,  $p < .001$ , Cohen’s  $d = .44$ ), this preference reversed (and was similar in magnitude) in the gain frame ( $b = -.30$ , 95% CI  $[-.38, -.22]$ ,  $t = 7.32$ ,  $p < .001$ , Cohen’s  $d = .34$ ).

To put these results in perspective, we assessed what percent of the time participants showed an overall preference for the risk-seeking or risk-averse candidate in each condition, defined as an overall leadership perception index greater than zero (indicating a preference for the risk-seeking candidate), an overall leadership perception index less than zero (indicating a preference for the risk-averse candidate), or an overall index equal to zero (indicating no preference). In the loss condition, participants showed a preference for the risk-seeking candidate 58% of the time compared with a preference for the risk-averse candidate just 21% of the time (the remaining 21% indicated no preference). In the gain condition, the pattern fully reversed: participants showed a preference for the risk-averse candidate 54% of the time compared with a preference for the risk-seeking candidate just 23% of the time (the remaining 23% indicated no preference).

**Figure 5**  
Preference for Risk-Seeking Candidate as a Function of Frame (Experiment 3,  $N = 906$ )



*Note.* In the gain condition, participants demonstrated a preference for the risk-averse candidate. This preference reversed in the loss condition. Violin plots display the distributions and horizontal bars indicate the means. Rectangles show 95% confidence intervals. See the online article for the color version of this figure.

Among participants who showed any preference (i.e., did not have an overall index equal to zero), participants preferred the risk-seeking candidate 73% of the time in the loss frame but just 30% of the time in the gain frame. These results persist despite the fact that the policies espoused by the candidates were identical across conditions.

Taken together, the results provided evidence that decision frames were socially relevant for leaders, even when the choice options were logically equivalent across frames.

### Individual Difference Analyses

In an exploratory analysis, we assessed whether three individual difference variables moderated the effect of the frame on perceived leadership effectiveness. We did not test any other moderators in this study, but full data and materials are available on Researchbox.com for interested readers.

First, we tested whether individuals higher in tendency for cognitive reflection would be less susceptible to changing their preferred politician based on the decision frame. Prior research suggests that individuals higher in tendency for cognitive reflection may be less susceptible to framing effects themselves (Frederick, 2005; Toplak, West, & Stanovich, 2011). To test this hypothesis, we regressed leadership preference on frame, tendency for cognitive reflection, and their interaction. We did not find evidence to support this hypothesis:  $b = .00$ , 95% CI  $[-.05, .06]$ ,  $t = .16$ ,  $p > .87$ . Second, we found a similar null result for self-reported knowledge of framing effects ( $b = -.06$ , 95% CI  $[-.32, .19]$ ,  $t = -.48$ ,  $p = .63$ ), in line with null results from Experiments 1 and 2. Finally, less than 5% of our sample indicated that they either were not registered to vote or did not plan to vote, so we did not conduct planned moderation analyses with these variables (though the results held when we excluded data from the small subset of participants who indicated they were not registered or did not plan to vote).

### Discussion

Experiment 3 replicated prior experiments with a new context, with joint evaluation, without first asking the participant's preference, and with a new participant sample.

### General Discussion

Loss-gain framing effects on risk preferences are among the most robust and influential findings identified in the past half century of social and behavioral science research (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981; for review, see Ruggeri et al., 2020). Traditionally, researchers have argued that the *economically* optimal choice never depends on the frame in which the options are presented (Bazerman & Moore, 2009; Tversky & Kahneman, 1981). In the present research, results reveal that the *reputationally* optimal choice sometimes does. Specifically, we hypothesized and found that while decision makers who made risk-averse (vs. risk-seeking) choices were socially and financially penalized by third-party observers when choice outcomes were framed as losses, these same risk-averse decision makers were socially rewarded when outcomes were framed as gains (and the reverse for risk-seeking decision makers). We found evidence for this pattern across five social perceptions, three decision contexts, two participant populations of United States adults, and with real monetary stakes in a dictator game.

We also provided evidence for a psychological process through which these effects occurred. Observers were more likely to

disagree with decision makers who made frame-inconsistent choices. In turn, disagreement was associated with negative reputational consequences for a wide variety of social perceptions, and decreased participants' willingness to share financial reward.

### Theoretical Implications

An influential and interdisciplinary body of research on human judgment and decision making contrasts prescriptive models of how individuals should behave with descriptive models of how individuals actually behave. Typically, deviations from prescriptive models are treated as suboptimal and in need of correction.<sup>4</sup> For example, in his Nobel Prize address, Kahneman (2002) summarized this body of research: "Our research attempted to obtain a map of bounded rationality, by exploring the systematic biases that separate the beliefs that people have and the choices they make from the optimal beliefs and choices assumed in rational-agent models."

But what makes an optimal belief or choice? Here, we argue that an expanded focus on the goals decision makers themselves hold (i.e., reputation management) questions whether such deviations from rational-agent models should always be considered suboptimal. We test this broader theorizing in the context of loss-gain framing effects on risk preferences not because we think the psychological dynamics at play are unique to this context, but rather because such framing effects have been uniquely influential for both academic discourse and applied interventions in policy and organizations. In fact, the results hold preliminary implications not only for decision making under risk but also for extending understanding of a range of other behavioral tendencies long considered irrational biases in the research literature on judgment and decision making (e.g., sunk cost bias; see Dorison et al., 2021).

An important clarification of our claims merits note. We are not claiming that it is always rational to be biased just because others are. This is especially true in cases where differences in nonreputational consequences are quite large and differences in reputational incentives are quite small. Of note, the vast majority of framing effect studies in the lab do not have an explicit reputational or strategic component at all, suggesting that attention to decision frames in these contexts should be considered irrational.

In other cases, however, the reputational benefits may outweigh the material costs to falling victim to apparently suboptimal behavior, such as framing effects. For example, in cases where differences in reputational consequences are quite large (e.g., for a public leader before an election or for an employee in the days before a big promotion decision at work) and differences in nonreputational consequences are quite small (e.g., choosing between two similar policies or options), we do suggest that decision makers may be justified (and thus rational) in attending to loss versus gain framing, given that they care deeply about their reputations. As such, we consider the present set of results an "existence proof" that qualifies the traditional view in the

<sup>4</sup> As an important exception, McKenzie and colleagues suggest that certain framing effects can be considered rational because the description of events implicitly conveys (or "leaks") relevant information, thus violating assumptions of informational equivalence across frames (McKenzie, 2004; see also McKenzie & Nelson, 2003; Sher & McKenzie, 2006, 2008).

literature that loss–gain framing effects on risk preferences are *always* irrational. Even if there are reputational advantages to a behavior, these reputational benefits must be balanced against material (i.e., nonreputational) costs.

Our primary focus was on the prescriptive level of analysis; however, our research also has preliminary theoretical implications for the descriptive level of analysis. Whereas prior work has traditionally explored cognitive explanations of framing effects (Tversky & Kahneman, 1981), the reputational incentives identified in this work may illuminate a novel driver for why framing effects (and other such deviations from prescriptive decision models) persist over time. If ignoring decision frames is reputationally costly, then individuals may be more likely to attend to them, even if decision makers are unaware of this social influence (for related work, see Jordan & Rand, 2020). Although the present work focused on third-party observers, it sets the foundation for future work examining to what extent, if at all, reputational incentives shift risk preferences for lay decision makers and leaders alike (for related work, see Lerner & Tetlock, 1999; Tetlock, 2000, 2002).

### Practical Implications

Our work also has potential practical applications for organizational behavior and leadership that merit further testing. First, the existence of such reputational costs suggests new potential levers to reduce framing effects—and a host of other behavioral biases. Whereas prior research has uncovered successful strategies for debiasing decision making by focusing on the cognition of the individual decision maker (Morewedge et al., 2015; Sellier et al., 2019), the present research suggests that an additional focus on the social environment (e.g., an organizational culture) could also serve as a successful strategy (Gigerenzer, 2014; Lerner & Tetlock, 1999). For example, a workplace program socially rewarding consistency in risk preferences could incentivize decision makers to ignore decision frames. Or, a culture of calculating expected value (agnostic to loss vs. gain framing) could encourage expected-value maximizing choices, irrespective of frames. An interesting related question is what strategies leaders can introduce to create a culture that values calculating expected values.<sup>5</sup> Future work is needed before any firm conclusions can be drawn about the potential effectiveness of developing such interventions and cultures.

Second, it could be the case that leaders in the public and private sector could harness decision frames to increase or decrease the social palatability of candidates or policies (Zlatev et al., 2017). This may already have happened. For example, as observed in an editorial in *The New York Times* (Vavreck, 2016) during the 2016 United States presidential election, Donald Trump—the candidate perceived as more of a risky option—typically focused on losses (e.g., jobs, greatness). On the other hand, Hillary Clinton—the candidate perceived as a less risky option—was more likely to focus on gains. In an organizational setting, our work suggests that leaders could strategically employ decision frames to build consensus and protect their reputations.

### Limitations and Future Directions

A few notable limitations merit note and thus provide fertile ground for future research. First, the present work examines the *reputational consequences* of ignoring decision frames, and thus

yields insights into the prescriptive claim about whether frames should always be ignored by decision makers. For this reason, the experiments focus on the behavior of third-party observers and are most centrally relevant to the question of what it means to make a good decision. Although related, the present article does not directly address the *reputational causes* of framing effects and thus does not directly test the descriptive claim that decision makers purposefully fall victim to framing effects in service of impressing observers. To test this claim, experiments should focus on the behavior of decision makers themselves, with the critical experimental manipulation testing (for example, whether they make different choices in public versus in private. These experiments would be most centrally relevant to the question of why decision makers fall victim to framing effects in the first place. All that being said, this distinction is a soft one, in which reputational causes and consequences influence each other in reciprocal ways, rather than a hard one, in which these are two completely unrelated research questions.

Second, ambiguity remains regarding the underlying mechanism, with important implications for boundary conditions of the present results. Our results make clear that ignoring decision frames heightens disagreement, which is associated with reputational costs. Important questions remain regarding precisely why disagreement is associated with reputational penalties and the boundary conditions for such effects. Our interpretation of our results is that such effects are likely to be symmetric across loss versus gain frames and broad-based in nature,<sup>6</sup> thus supporting an interpretation based on halo effects (rather than the less parsimonious naïve realism). However, future research is needed to more rigorously tease apart predictions based on halo effects versus naïve realism by assessing other dependent variables, which might predict that some social perceptions (e.g., perceptions of bias/competence) would be affected more than others (e.g., perceptions of warmth). Additionally, future research could manipulate whether observers indicate their preferred option before versus after reading about the other person's choice (rather than just eliminate it, as in Experiment 3).

Third, and related, future research is needed to assess a broader set of outcome variables to assess whether such results should be considered penalties for disagreement or reward for agreement. The data show a clear differentiation between social perceptions of (or financial rewards to) decision makers who make frame-consistent versus frame-inconsistent choices. However, it is much less clear whether this differentiation should be attributed to rewarding the frame-consistent decision maker or penalizing the frame-inconsistent decision maker, as both interpretations could be considered equally valid from the differentiation between the two.

<sup>5</sup> For example, the so-called “analytics revolution” in professional sports may embody this type of a cultural shift, inducing a concomitant change in reputational incentives that explains changes in risk attitudes among players or coaches. For example, future research could assess whether the documented rise in rates of attempting risky fourth down conversions among NFL teams is associated with changes in the way the potential or realized outcomes of such attempts are discussed in the media or by team personnel.

<sup>6</sup> As an important exception, there was virtually no effect on perceptions of competence in the gain frame. We do not have any explanation for this asymmetry, although believe it is an exciting avenue for future research.

The data from Experiment 2 provide some suggestive evidence for a penalty-focused interpretation. In that experiment, the slider used to measure dictator game giving was defaulted to 50% of the allocation. Given that average dictator game giving in all conditions was less than 50% of the allocation, this appears to fit an interpretation that the average observer chose to take additional money away from the other person (in the form of a penalty) compared with the default allocation. Future research is needed to more rigorously test whether the set of results should be considered reward versus penalties. One possibility is to do so through other financially-incentivized behavioral-economic games, such as a trust or distrust game (e.g., Berg et al., 1995; Everett et al., 2021; McEvily et al., 2012).

Fourth, an additional methodological question is whether our results would hold in a within- (rather than between-) subjects design. In our work, we designed our experiments (i.e., the between-subjects design) to mirror the situations in which people typically find themselves in the “real world,” in which they evaluate choices made in a single frame. Rarely (at least in our experience) do people naturally undertake the counterfactual to think about the opposite frame. However, an important open question is whether decision makers would be rewarded for consistency if observers evaluated their choices in both frames simultaneously. Indeed, when people learn about framing effects, a common reaction (although not one that the authors themselves endorse) is that the violation of transitivity in risk preferences shows that people are “stupid.” Our theory would predict that if confronting observers with both frames was effective in changing observers’ own internal preferences, then it could be sufficient to influence their external evaluations, as well. That being said, participants’ self-reported exposure to the principle of framing effects (collected as an exploratory individual difference variable at the end of all experiments) did not significantly moderate the results in any of the three experiments, suggesting that it may be necessary not only to teach people about framing effects in general, but also to make both frames explicitly salient at the time of evaluation to attenuate differences in social perceptions.

Finally, whereas the present work used two different online samples of adults in the United States, it could be the case that certain populations, such as those trained in the social and behavioral sciences or from different cultures, would *not* show the same pattern of results. That being said, results in our experiments held across age, gender, education, self-reported knowledge of framing effects, and tendency toward cognitive reflection.

## Conclusion

A mature literature on human judgment and decision making documents deviations between prescriptive and descriptive models, with the effect of loss–gain framing on risk preferences as perhaps the most prominent example (Tversky & Kahneman, 1981; for review, see Ruggeri et al., 2020). Traditionally, deviations such as framing effects are considered suboptimal and in need of correction. In the present research, we provide evidence that decision frames can be reputationally relevant at the same time they are economically irrelevant. Results suggest that this occurs because third-party observers themselves are influenced by decision frames and punish disagreement. Given that decision makers hold deep impression management concerns, such reputational consequences qualify the strong prescriptive that decision makers should always remain

immune to framing effects. Future research is needed to explore boundary conditions in both the lab and field, examine whether results generalize to other behavioral biases (e.g., sunk cost bias), and test interventions based on understanding the reputational costs of economic rationality.

## Context

This article is part of a broader program of research that reconsiders classic work on decision making biases through a social/organizational lens, revealing that several seemingly suboptimal behaviors appear functional once reputational incentives are introduced. This program of research arose from discussions with public leaders in executive education programs, who argued (often quite vehemently) that our decision models were incomplete and did not take into account key relational incentives that were central to their ability to successfully lead their teams and organizations. We are extremely grateful for their expertise and insights in helping us develop and refine our decision models and thus hopefully better serve them as they work tirelessly to improve health and well-being for citizens globally.

## References

- Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology, 49*(3), 323–330. <https://doi.org/10.1037/0022-006X.49.3.323>
- Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110*(3), 486–498. <https://doi.org/10.1037/0033-2909.110.3.486>
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529. <https://doi.org/10.1037/0033-2909.117.3.497>
- Bazerman, M. H. (1983). Negotiator judgment: A critical look at the rationality assumption. *American Behavioral Scientist, 27*(2), 211–228. <https://doi.org/10.1177/000276483027002007>
- Bazerman, M. H., & Moore, D. A. (2009). *Judgment in managerial decision making*. Wiley.
- Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics, 110*(1), 73–92. <https://doi.org/10.2307/2118511>
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142. <https://doi.org/10.1006/game.1995.1027>
- Briggs, R. A. (2019). Normative theories of rational choice: Expected utility. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/fall2019/entries/rationalitynormative-utility/>
- Byrne, D. (1961). Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology, 62*, 713–715. <https://doi.org/10.1037/h0044721>
- Chen, M. K., Lakshminarayanan, V., & Santos, L. R. (2006). How basic are behavioral biases? Evidence from capuchin monkey trading behavior. *Journal of Political Economy, 114*(3), 517–537. <https://doi.org/10.1086/503550>
- Dehghani, M., Johnson, K., Hoover, J., Sagi, E., Garten, J., Parmar, N. J., Vaisey, S., Iliev, R., & Graham, J. (2016). Purity homophily in social networks. *Journal of Experimental Psychology: General, 145*(3), 366–375. <https://doi.org/10.1037/xge0000139>
- DiPrete, T. A., Gelman, A., McCormick, T., Teitler, J., & Zheng, T. (2011). Segregation in social networks based on acquaintanceship and trust. *American Journal of Sociology, 116*(4), 1234–1283. <https://doi.org/10.1086/659100>

- Dorison, C. A., Umphres, C. K., & Lerner, J. S. (2021). Staying the course: Decision makers who escalate commitment are trusted and trustworthy. *Journal of Experimental Psychology: General*. Advance online publication. <https://doi.org/10.1037/xge0001101>
- Everett, J. A. C., Colombatto, C., Awad, E., Boggio, P., Bos, B., Brady, W. J., Chawla, M., Chituc, V., Chung, D., Drupp, M. A., Goel, S., Grosskopf, B., Hjorth, F., Ji, A., Kealoha, C., Kim, J. S., Lin, Y., Ma, Y., Maréchal, M. A., . . . Crockett, M. J. (2021). Moral dilemmas and trust in leaders during a global health crisis. *Nature Human Behaviour*, 5(8), 1074–1088. <https://doi.org/10.1038/s41562-021-01156-y>
- Everett, J. A. C., Faber, N. S., Savulescu, J., & Crockett, M. J. (2018). The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *Journal of Experimental Social Psychology*, 79, 200–216. <https://doi.org/10.1016/j.jesp.2018.07.004>
- Fagley, N. S. (1993). A note concerning reflection effects versus framing effects. *Psychological Bulletin*, 113(3), 451–452. <https://doi.org/10.1037/0033-2909.113.3.451>
- Fast, N. J., Sivanathan, N., Mayer, N. D., & Galinsky, A. D. (2012). Power and overconfident decision-making. *Organizational Behavior and Human Decision Processes*, 117(2), 249–260. <https://doi.org/10.1016/j.obhdp.2011.11.009>
- Fischhoff, B., & Broomell, S. B. (2020). Judgment and decision making. *Annual Review of Psychology*, 71, 331–355. <https://doi.org/10.1146/annurev-psych-010419-050747>
- Fiske, S. T. (2015). Intergroup biases: A focus on stereotype content. *Current Opinion in Behavioral Sciences*, 3, 45–50. <https://doi.org/10.1016/j.cobeha.2015.01.010>
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-based trust in strangers: The role of stereotypes and expectations. *Psychological Science*, 20(4), 419–422. <https://doi.org/10.1111/j.1467-9280.2009.02312.x>
- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>
- Gigerenzer, G. (2014). Should patients listen to how doctors frame messages? *BMJ (Clinical Research Ed.)*, 349, g7091. <https://doi.org/10.1136/bmj.g7091>
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–482. <https://doi.org/10.1146/annurev-psych-120709-145346>
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103(4), 650–669. <https://doi.org/10.1037/0033-295X.103.4.650>
- Gilovich, T. D., & Griffin, D. W. (2010). Judgment and decision making. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th edition, pp. 542–588). John Wiley.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511808098>
- Goffman, E. (1959). *The presentation of self in everyday life*. Harmondsworth.
- Grossmann, I., Eibach, R. P., Koyama, J., & Sahi, Q. B. (2020). Folk standards of sound judgment: Rationality versus reasonableness. *Science Advances*, 6(2), eaaz0289. <https://doi.org/10.1126/sciadv.aaz0289>
- Halberstam, Y., & Knight, B. (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, 143, 73–88. <https://doi.org/10.1016/j.jpubeco.2016.08.011>
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.
- Hershey, J. C., & Schoemaker, P. J. (1980). Risk taking and problem context in the domain of losses: An expected utility analysis. *The Journal of Risk and Insurance*, 47(1), 111–132. <https://doi.org/10.2307/252685>
- Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, 125(5), 576–590. <https://doi.org/10.1037/0033-2909.125.5.576>
- Huber, G. A., & Malhotra, N. (2017). Political homophily in social relationships: Evidence from online dating behavior. *The Journal of Politics*, 79(1), 269–283. <https://doi.org/10.1086/687533>
- John, L. K., Jeong, M., Gino, F., & Huang, L. (2019). The self-presentational consequences of upholding one's stance in spite of the evidence. *Organizational Behavior and Human Decision Processes*, 154, 1–14. <https://doi.org/10.1016/j.obhdp.2019.07.001>
- Jordan, J. J., & Rand, D. G. (2020). Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *Journal of Personality and Social Psychology*, 118(1), 57.
- Jordan, J. J., Hoffman, M., Nowak, M. A., & Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences of the United States of America*, 113(31), 8658–8663. <https://doi.org/10.1073/pnas.1601280113>
- Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. *Nobel Prize Lecture*, 8, 351–401.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1), 193–206. <https://doi.org/10.1257/jep.5.1.193>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291. <https://doi.org/10.2307/1914185>
- Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2), 405–450. <https://doi.org/10.1086/599247>
- Kreps, T. A., Laurin, K., & Merritt, A. C. (2017). Hypocritical flip-flop, or courageous evolution? When leaders change their moral minds. *Journal of Personality and Social Psychology*, 113(5), 730–752. <https://doi.org/10.1037/pspi0000103>
- Leary, M. R., Raimi, K. T., Jongman-Sereno, K. P., & Diebels, K. J. (2015). Distinguishing intrapsychic from interpersonal motives in psychological theory and research. *Perspectives on Psychological Science*, 10(4), 497–517. <https://doi.org/10.1177/1745691615583132>
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255–275. <https://doi.org/10.1037/0033-2909.125.2.255>
- Lieberman, V., Minson, J. A., Bryan, C. J., & Ross, L. (2012). Naïve realism and capturing the “wisdom of dyads.” *Journal of Experimental Social Psychology*, 48(2), 507–512. <https://doi.org/10.1016/j.jesp.2011.10.016>
- Loewenstein, G., Rick, S., & Cohen, J. D. (2008). Neuroeconomics. *Annual Review of Psychology*, 59, 647–672. <https://doi.org/10.1146/annurev.psych.59.103006.093710>
- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127(2), 267–286. <https://doi.org/10.1037/0033-2909.127.2.267>
- McEvily, B., Radzevick, J. R., & Weber, R. A. (2012). Whom do you distrust and how much does it cost? An experiment on the measurement of trust. *Games and Economic Behavior*, 74(1), 285–298. <https://doi.org/10.1016/j.geb.2011.06.011>
- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, 111(2), 361–365. <https://doi.org/10.1037/0033-2909.111.2.361>
- McKenzie, C. R. (2004). Framing effects in inference tasks—And why they are normatively defensible. *Memory & Cognition*, 32(6), 874–885. <https://doi.org/10.3758/BF03196866>

- McKenzie, C. R., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review*, *10*(3), 596–602. <https://doi.org/10.3758/BF03196520>
- McNeil, B. J., Pauker, S. G., Sox, H. C., Jr., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *The New England Journal of Medicine*, *306*(21), 1259–1262. <https://doi.org/10.1056/NEJM198205273062103>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Mellers, B. A., Schwartz, A., & Cooke, A. D. (1998). Judgment and decision making. *Annual Review of Psychology*, *49*(1), 447–477. <https://doi.org/10.1146/annurev.psych.49.1.447>
- Minson, J. A., & Dorison, C. A. (2021). Toward a psychology of attitude conflict. *Current Opinion in Psychology*, *43*, 182–188. <https://doi.org/10.1016/j.copsyc.2021.07.002>
- Minson, J. A., & Mueller, J. S. (2012). The cost of collaboration: Why joint decision making exacerbates rejection of outside information. *Psychological Science*, *23*(3), 219–224. <https://doi.org/10.1177/09567971611429132>
- Minson, J. A., Liberman, V., & Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality and Social Psychology Bulletin*, *37*(10), 1325–1338. <https://doi.org/10.1177/0146167211410436>
- Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129–140. <https://doi.org/10.1177/2372732215600886>
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. <https://doi.org/10.1037/0033-295X.84.3.231>
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*(1), 87–131. <https://doi.org/10.1146/annurev.ps.43.020192.000511>
- Plato. (1968). *Laws. Plato in twelve volumes* (Vol. 11). Harvard University Press.
- Prinstein, M. J., & Dodge, K. A. (Eds.). (2008). *Understanding peer influence in children and adolescents*. Guilford Press.
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, *111*(3), 781–799. <https://doi.org/10.1037/0033-295X.111.3.781>
- Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science*, *325*(5945), 1272–1275. <https://doi.org/10.1126/science.1177418>
- Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: “Naive realism” in intergroup perception and conflict. *Journal of Personality and Social Psychology*, *68*(3), 404–417. <https://doi.org/10.1037/0022-3514.68.3.404>
- Ross, L. (2018). From the fundamental attribution error to the truly fundamental attribution error and beyond: My research journey. *Perspectives on Psychological Science*, *13*(6), 750–769. <https://doi.org/10.1177/1745691618769855>
- Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. *Advances in Experimental Social Psychology*, *27*, 255–304.
- Rossee, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA.). *Journal of Statistical Software*, *48*(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Ruggeri, K., Alf, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., Davison, C., Demic, E., Esteban-Serna, C., Friedemann, M., Gibson, S. P., Jarke, H., Karakasheva, R., Khorrami, P. R., Kveder, J., Anderson, T. L., Lofthus, I. S., McGill, L., Nieto, A. E., . . . Folke, T. (2020). Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, *4*, 622–633. <https://doi.org/10.1038/s41562-020-0886-x>
- Schlenker, B. R., & Weigold, M. F. (1992). Interpersonal processes involving impression regulation and management. *Annual Review of Psychology*, *43*(1), 133–168. <https://doi.org/10.1146/annurev.ps.43.020192.001025>
- Sellier, A. L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychological Science*, *30*(9), 1371–1379. <https://doi.org/10.1177/0956797619861429>
- 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>
- Sher, S., & McKenzie, C. R. M. (2008). Framing effects and rationality. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science*, (pp. 79–96). Oxford University Press.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. <http://ssrn.com/abstract=2160588> or <https://doi.org/10.2139/ssrn.2160588>
- Sun, Y., & Mellers, B. (2016). Trade-upgrade framing effects: Trades are losses, but upgrades are improvements. *Judgment and Decision Making*, *11*(6), 582–588.
- Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–102. <https://doi.org/10.1038/scientificamerican1170-96>
- Tenney, E. R., Meikle, N. L., Hunsaker, D., Moore, D. A., & Anderson, C. (2019). Is overconfidence a social liability? The effect of verbal versus nonverbal expressions of confidence. *Journal of Personality and Social Psychology*, *116*(3), 396–415. <https://doi.org/10.1037/pspi0000150>
- Tetlock, P. E. (2000). Cognitive biases and organizational correctives: Do both disease and cure depend on the politics of the beholder? *Administrative Science Quarterly*, *45*(2), 293–326. <https://doi.org/10.2307/2667073>
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review*, *109*(3), 451–471. <https://doi.org/10.1037/0033-295X.109.3.451>
- Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99.
- 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>
- 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>
- Vavreck, L. (2016, May 9). Big picture for Trump vs. Clinton? It's all in the framing. *The New York Times*. <http://www.nytimes.com>
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.
- Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, *60*, 53–85. <https://doi.org/10.1146/annurev.psych.60.110707.163633>
- Zlatev, J. J., Daniels, D. P., Kim, H., & Neale, M. A. (2017). Default neglect in attempts at social influence. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(52), 13643–13648. <https://doi.org/10.1073/pnas.1712757114>

Received May 18, 2021

Revision received December 16, 2021

Accepted December 16, 2021 ■