

**Third-party observers socially and financially penalize decision makers who do not
fall victim to framing effects**

Charles A. Dorison¹ and Blake H. Heller²

¹Kellogg School of Management, Northwestern University

²Harvard Kennedy School, Harvard University

Author Note

All data, study materials, analysis code, and pre-registrations are publicly available [here](#).

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 manuscript. The authors declare no competing interests or conflicts.

Correspondence concerning this article should be addressed to Charles A. Dorison, Northwestern University. Email: charles.dorison@kellogg.northwestern.edu

Word count: 7,732

Abstract

A large interdisciplinary body of research on human judgment and decision making documents systematic deviations between prescriptive models (i.e., how individuals *should* behave) and descriptive models (i.e., how individuals *actually* behave). One canonical example of such deviations is the framing effect: the robust tendency for risk preferences to shift depending on whether outcomes are described as gains or losses. While prior research argues that rational decision makers should be immune to framing effects, we present three pre-registered experiments (N=1,954) that qualify this prescription. We predict and find that while third-party observers penalize decision makers who make risk-averse (versus 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. In line with research on naïve realism, this pattern is driven by the fact that observers themselves fell victim to framing effects and punished decision makers who disagreed. 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 can ultimately reveal not only why such biases are so prevalent but also novel interventions. Specifically, our results suggest a necessary focus on the social environment in addition to the cognition of the decision maker.

Keywords: Decision Making, Framing, Risk Taking, Reputation, Naïve Realism

Drawing on research in psychology, economics, and neuroscience, over 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, Griffin, & Kahneman, 2002; Kahneman, Knetsch, & Thaler, 1991; Loewenstein, Rick, & Cohen, 2008; Loewenstein, Weber, Hsee, & Welch, 2001; Mellers, Schwartz, & Cooke, 1998; Payne, Bettman, & Johnson, 1992; Weber & Johnson, 2009). One canonical example of these deviations is the framing effect: the robust tendency for risk preferences to shift depending on whether outcomes are described as gains or losses (Kahneman & Tversky, 1979; Tversky & Kahneman 1981). Specifically, individuals are more likely to make risk-averse choices when the expected outcomes of alternative choices are framed as gains and are more likely to make risk-seeking choices when outcomes are framed as losses, even when choice sets are otherwise equivalent.

Researchers across disciplines have documented 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, 1985), reaching agreement in negotiations (Bazerman, 1983), and trading commodities (Sun & Mellers, 2016). Framing effects generalize in samples ranging from professional physicians (McNeil, Pauker, Sox, & Tversky, 1982) to capuchin monkeys (Chen, Lakshminarayanan, & Santos, 2006). More recently, an international project successfully replicated framing effects across 19 countries and 13

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

Traditionally, decision frames are treated as irrelevant features of a decision that *should* be ignored when making choices—even though they are influential in practice. 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.” In a 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.”

Present aims

In the present work, we qualify this strong prescriptive claim by documenting the reputational costs of ignoring decision frames. We hypothesize that while ignoring decision frames is rational under the parsimonious assumptions of traditional choice models (Briggs, 2019; von Neumann & Morgenstern, 1944), even rational decision makers can be affected by decision frames because frames change their reputational incentives. We test our 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.

Given that decision makers often hold deep impression management concerns (Lerner & Tetlock, 1999; Schlenker & Weigold, 1992; Tetlock, 2002), our results challenge the long-standing prescription that they should always ignore decision frames. Further, they hold implications not only for framing effects, but also for extending understanding of a range of behavioral tendencies long considered irrational biases. Such understanding can ultimately reveal not only why such biases are so prevalent but also novel interventions. Specifically, our results suggest a necessary focus on the social environment in addition to the cognition of the decision maker. We discuss these implications further in the General Discussion.

Theoretical Background

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

Second, we draw on theory and research on naïve realism (Pronin, Gilovich, & Ross, 2004; Robinson, Keltner, Ward, & Ross, 1995; Ross & Ward, 1995) to predict that disagreement will in turn drive reputational penalties. Prior research on naïve realism, sometimes called the “objectivity illusion” (Ross, 2018), 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, Gilovich, & Ross, 2004; Ross, 2018).

While initial evidence for the naïve realism hypothesis was rooted in the domains of partisan conflict and negotiations (e.g., Robinson, Keltner, Ward, & Ross, 1995), researchers testing this hypothesis have also branched out to domains in which individuals do not hold strong, identity-relevant attitudes. For example, Minson and colleagues (Minson, Liberman, & Ross, 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, Minson, Bryan, & Ross, 2012; Minson & Mueller, 2012).

We predicted that in the domain of risk preferences, disagreement over the “right” decision would underpin (i.e., statistically mediate) negative inferences. Notably, these findings reinforce longstanding psychological theories demonstrating preferences for similar over dissimilar others (e.g., Byrne, 1961; Tajfel, 1970).¹

Research Overview

We conducted three pre-registered experiments (collective N = 1,954) to test the hypothesis that while individuals who make risk-averse (versus risk-seeking) choices will

¹ They also converge with ancient philosophy: for example, Plato observed “Similarity begets friendship” and Aristotle noted “People love those who are like themselves.”

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. Our main claim is *not* the descriptive assertion that decision makers purposefully fall victim to framing effects in order to make a positive impression on observers. Rather, 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 by third-party observers rather than choices by decision makers themselves.

Open science statement. We report how we determined our sample size, all exclusions, all manipulations, and all measures in all experiments (Simmons, Nelson, & Simonsohn, 2012). All experiments were pre-registered on aspredicted.org and all pre-registrations, materials, data, and code are available on researchbox.org.²

Experiment 1

Experiment 1 tested whether ignoring decision frames carries reputational costs in the canonical influenza problem (Tversky & Kahneman, 1981, Problems 1 & 2). Participants in this experiment played the role of third-party 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) x 2 (Target choice: risk-averse, risk-seeking) design, as described below.

² Researchbox: 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>

Participants. Experiment 1 was conducted in November 2019, before the COVID-19 pandemic. 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 0.30 – 0.50) for simple effects. Due to chance factors, we ended up with 448 participants (199 women; mean age = 36.90, age range = 19 - 70). In accordance with our pre-registration, 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, while 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. While 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: perceived competence, warmth, and confidence. Order of presentation for the three variables was randomized. To measure competence and warmth (Fiske, Cuddy, Glick, & Xu, 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, Sivanathan, Mayer, & Galinsky, 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: perceived morality and leadership effectiveness. 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 ($b = -0.62$, 95% CI = [-1.02, -0.22], $t = -3.08$, $p = .002$), warmth ($b = -0.77$, 95% CI = [-1.16, -0.37], $t = -3.82$, $p < .001$), and confidence ($b = -0.46$, 95% CI = [-0.87, -0.06], $t = -2.28$, $p = .023$). We also found a significant interaction for our two exploratory dependent variables: perceptions of morality ($b = -0.73$, 95% CI = [-1.13, -0.33], $t = -3.62$, $p < .001$) and leadership effectiveness ($b = -0.86$, 95% CI = [-1.26, -0.47], $t = -4.23$, $p < .001$). 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-seeking target was directionally preferred in the loss frame. Further results are presented in Table 1.

Taken together, this set of results demonstrate 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.

Figure 1. Interaction between frame (loss, gain) and partner choice (risk-seeking, risk-averse) on five different social perceptions in Experiment 1 (N = 381). Black bars are evaluations of a risk-averse target and 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. While risk-seeking choices were socially rewarded when options were framed as losses, risk-averse choices were socially rewarded when options were framed as gains.

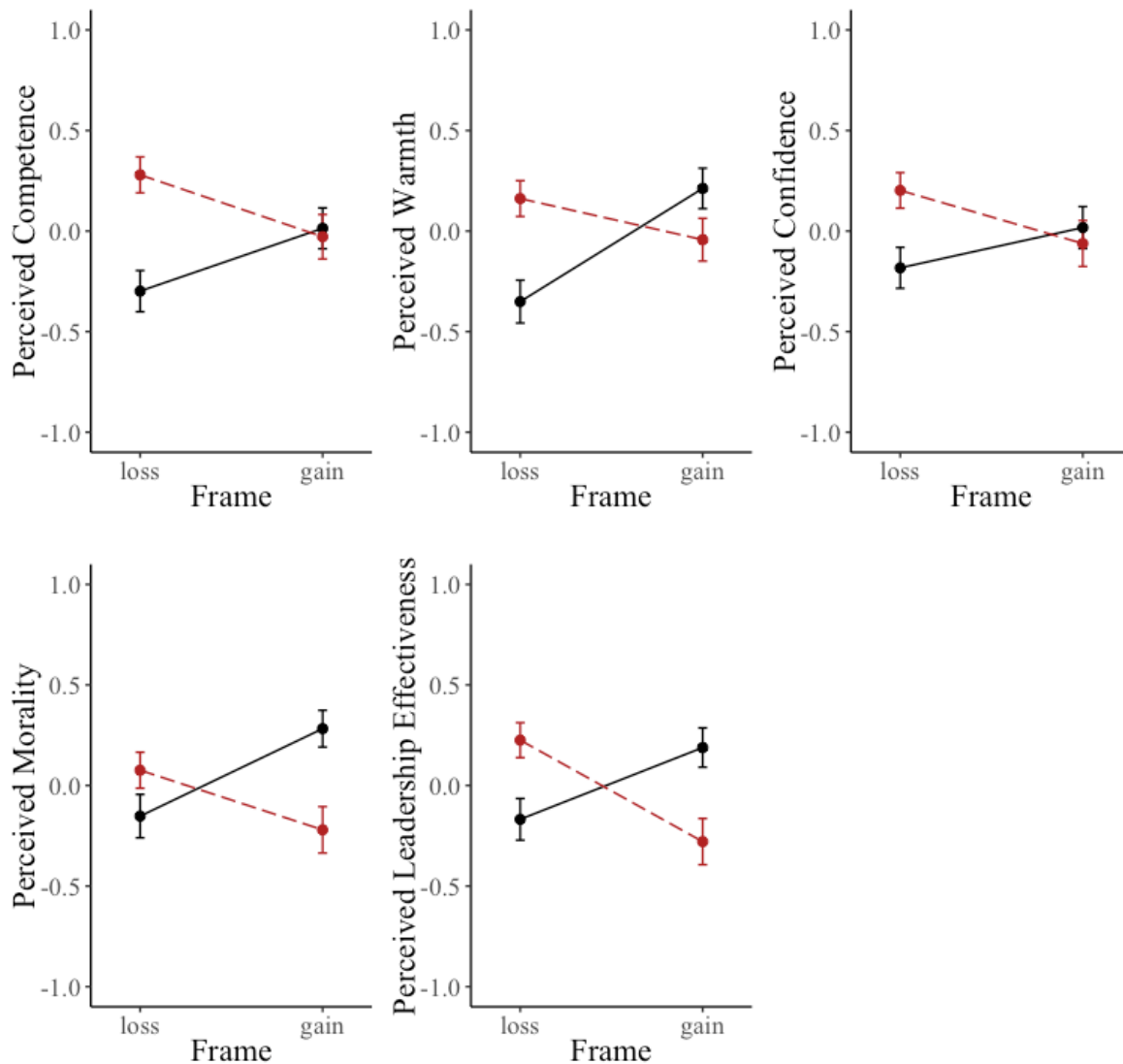


Table 1. Regression results for five interaction models in Experiment 1 (N = 381). The dependent variable was standardized in all cases. In all five models, there was a significant interaction between frame (loss, gain) and partner choice (risk-seeking, risk-averse).

	<i>Dependent variable:</i>				
	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 x 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 ²	0.043	0.049	0.020	0.048	0.038

*p<0.05; **p<0.01; ***p<0.001

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 reinforce the consistent reputational rewards associated with making frame-consistent choices, despite the logical equivalence of the choices across frames.

Mediation analyses. A key remaining question is whether agreement with the choice of the decision maker underpinned the interaction we observed regarding social perceptions, as predicted by theory and research on naïve realism. To answer this question, we fit five separate moderated mediation models using the Lavaan package in R (Rosseel, 2012). The models are 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 = -0.27, -0.29, -0.21, -0.34, and -0.42 respectively, all $ps < .003$). First, we found a significant interaction between the target decision maker's choice and frame on agreement ($b = -0.52$, 95% CI = [-0.71, -0.33], $t = -5.34$, $p < .001$, identical in all five models). This occurred because observers themselves fell victim to the framing effect: 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 ($bs = 0.51, 0.56, 0.41, 0.66, \text{ and } 0.81$ respectively, all $ps < .001$).

Taken together, across the five moderated mediation models, the results provide evidence consistent with the hypothesis that agreement with the target decision maker's choice itself underpinned social perceptions. These results are consistent with the theory that observers serve as naïve realists, socially derogating decision makers with diverging preferences.

Figure 2. One of five moderated mediation models fit in Experiment 1 (N = 381). 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.

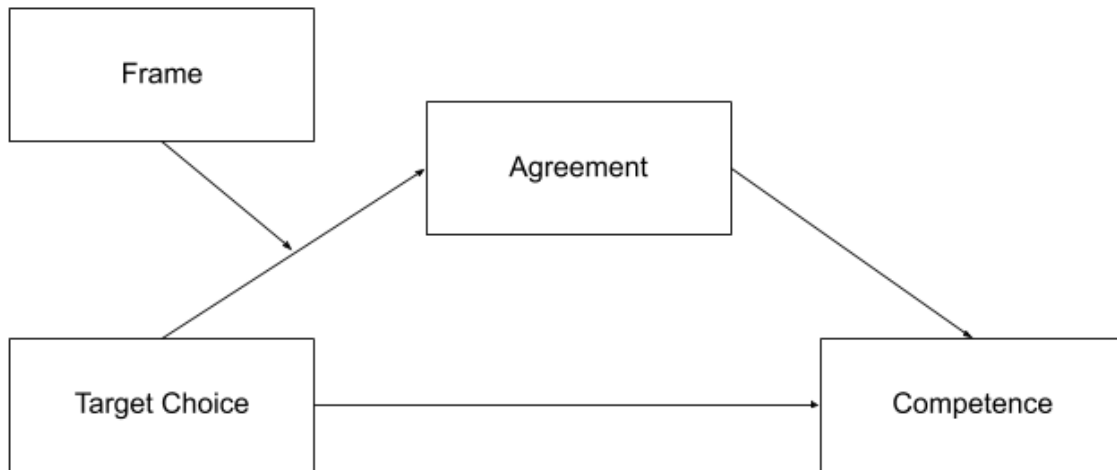


Table 2. Results of moderated mediation analyses in Experiment 1 (N = 381). In all five models, the independent variable was target choice, the mediating variable was agreement, and the moderating variable was frame. The dependent variables were standardized competence, warmth, confidence, morality, and leadership effectiveness in Models 1-5, respectively.

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%

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 ($b = -0.37$, 95% CI = [-0.71, -0.03], $t = -2.13$, $p = .034$). This result does not hold after accounting for multiple hypothesis testing and is likely due to chance 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 suggest that our results are not driven by any particular subpopulation and are broadly generalizable among the subgroups we consider.

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 tests the same overarching hypotheses with another canonical scenario and with real financial stakes. Experiment 2 extends Experiment 1 in at least two 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 are limited to life-or-death decisions or whether the pattern of results from Experiment 1 generalizes beyond the specific scenario considered therein. Second, we now implemented a dictator game with real financial stakes, in addition to social judgments. We predicted that, even with a new scenario and real money at stake, 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 rewards.

Method

As in Experiment 1, participants were randomly assigned to one of four between-subjects experimental conditions in a 2 (Frame: Loss, Gain) x 2 (Partner preference: risk-averse, risk-seeking) design. Participants again played the role of third-party observer. The scenario, described in greater detail below, was an adaptation of Problems 3 & 4 from Tversky & 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." Sample size again yielded 80% power to detect moderate effect sizes (i.e., Cohen's d between 0.30 – 0.50) for simple effects. Due to chance factors, we ended up

with 444 participants (161 women; mean age = 35.53, age range = 19 - 72). In accordance with our pre-registration, 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 \$0.20 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 actually 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 \$1000 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 \$1000 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.

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 (\$0.20) shared with the (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 amount sent to a partner ($b = -0.45$, 95% CI = [-0.86, -0.04], $t = -2.18$, $p = .030$). For ease of interpretation, we also report the impact of frame and partner choice on the non-standardized percentage of endowment sent to a partner ($b = -15.2\%$, 95% CI = [-34.6%, -5.4]) and the number of cents (out of a maximum of 20) sent to a partner ($b = -3.04$, 95% CI = [-5.91, -0.84]).

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 ($b = -0.72$, 95% CI = [-1.13, -0.32], $t = -3.55$, $p < .001$) and warmth ($b = -0.44$, 95% CI = [-0.84, -0.03], $t = -2.13$, $p = .034$). Results are visualized in Figure 3 and presented in Table 3.

Figure 3. Interaction between frame (loss, gain) and partner choice (risk-seeking, risk-averse) on three different dependent variables in Experiment 2 (N = 369). Black bars are evaluations of a risk-averse target and 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.

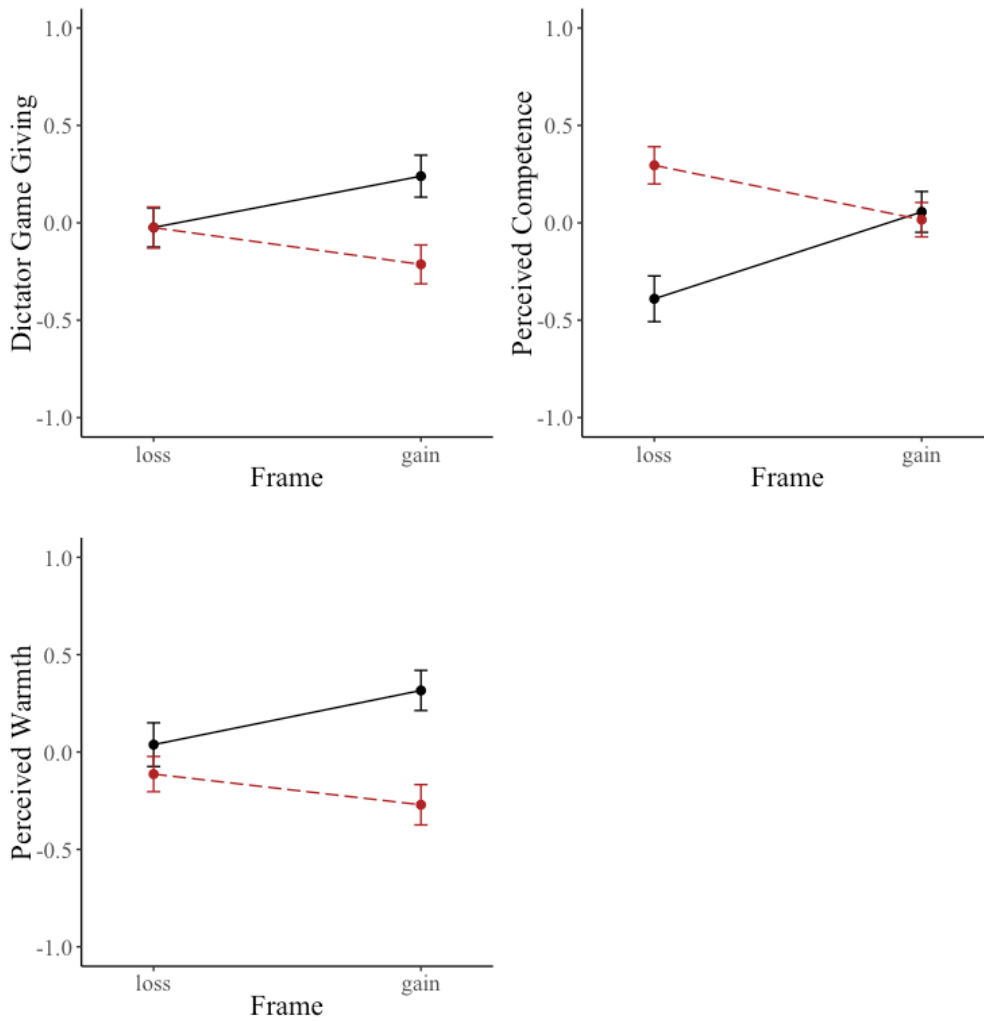


Table 3. Regression results for three interaction models in Experiment 2 (N = 369). The dependent variable was standardized in all cases. In all three models, there was a significant interaction between frame (loss, gain) and partner choice (risk-seeking, risk-averse).

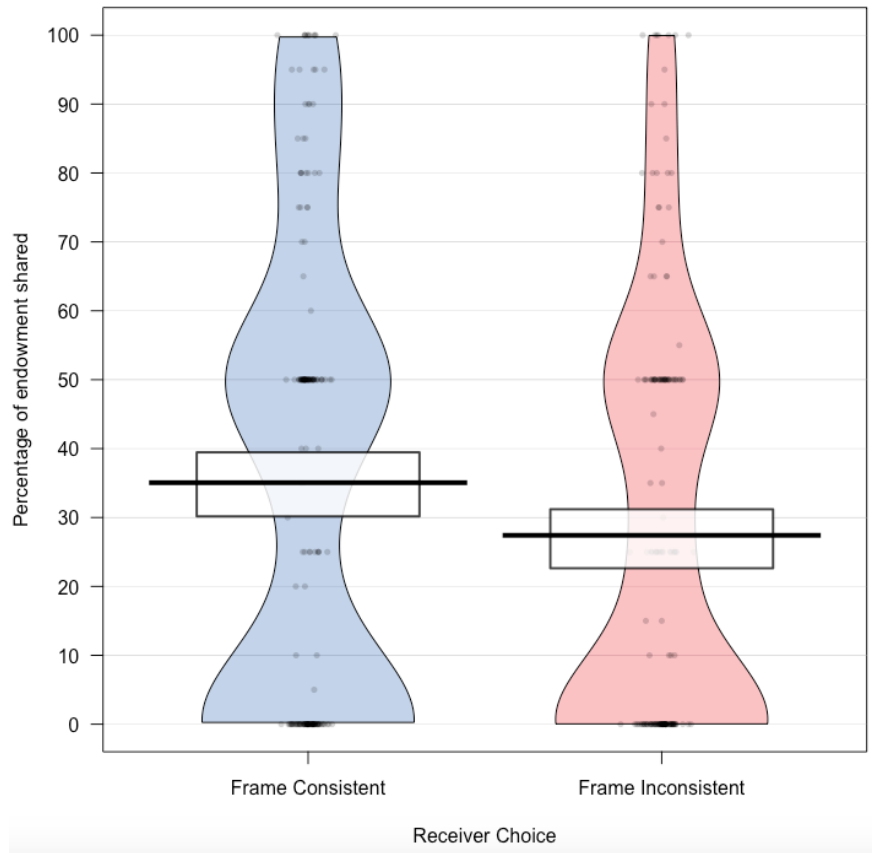
	<i>Dependent variable:</i>		
	dictator game giving	competence	warmth
	(1)	(2)	(3)
Risky choice	0.263 (0.146)	0.446** (0.144)	0.278 (0.144)
Gain frame	-0.001 (0.149)	0.685*** (0.147)	-0.151 (0.148)
Choice x Frame	-0.452* (0.208)	-0.725*** (0.204)	-0.436* (0.205)
Constant	-0.024 (0.107)	-0.390*** (0.105)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

*p<0.05; **p<0.01; ***p<0.001

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 to partners who made a frame-consistent choice (i.e., risk-seeking in the loss frame or risk-averse in the gain frame). While 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 = [0.94, 14.34], $t = 2.24$, $p = .026$) and was indicative of a relative increase of

almost 30% (an approximately 8% increase on a 27% baseline). Results are presented in Figure 4.

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 to receivers who made frame-inconsistent choices (risk-averse in the loss frame or risk-seeking in the gain frame) (Experiment 2, N = 369). Violin plots display the distributions and horizontal bars indicate the means. Rectangles show 95% confidence intervals.



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 (1) they differentiated

between partners on perceptions of warmth in the gain frame (but not the loss frame) and (2) 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 rewards 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 = -0.23, 95% CI = [-0.44, -0.02], $t = -2.09$, $p = .037$). First, we find a significant interaction between target choice and frame on agreement ($b = -1.00$, 95% CI = [-1.16, -0.83], $t = -11.79$, $p < .001$). 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, while observers in the gain frame preferred the risk-seeking choice only 8% of the time. Second, agreement positively predicted dictator game giving ($b = 0.23$, 95% CI = [0.02, 0.43], $t = 2.17$, $p = .030$). 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 = -0.72 and -0.37, 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 provide evidence consistent with the

hypothesis that agreement with a target decision maker’s choice underpinned social and financial rewards from third-party observers.

Table 4. Results of moderated mediation analyses in Experiment 2 (N = 369). 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.

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%

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 suggest that our results are not driven by any particular subpopulation and are broadly generalizable among the subgroups we consider.

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.

Experiment 3

Experiment 3 addresses four further potential theoretical and methodological boundary conditions of Experiments 1 and 2. First, we tested the same overarching hypothesis in the context of an unfolding public health crisis (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, Loewenstein, Blount, & Bazerman, 1999). Third, while 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 third-party observer. The scenario was an adaptation of Problems 1 & 2 from Tversky & 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 pre-registered 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 pre-registration, 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 0.20) 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 mid-sized 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 (1) the order of the risk-averse and risk-seeking candidates and (2) the names of the candidates were counterbalanced.

Social perceptions. We collected five items measuring perceived leadership ability of the candidates for governor. Participants indicated (1) which candidate they thought was a better leader; (2) which candidate they would vote for; (3) which candidate they thought would be a more effective governor; (4) which candidate they thought had better policies; and (5) 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.”³ Scores were re-coded 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.

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 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

³ In our pre-registration, we had a typo that indicated that these questions were answered on a 7-point scale.

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 includes 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.

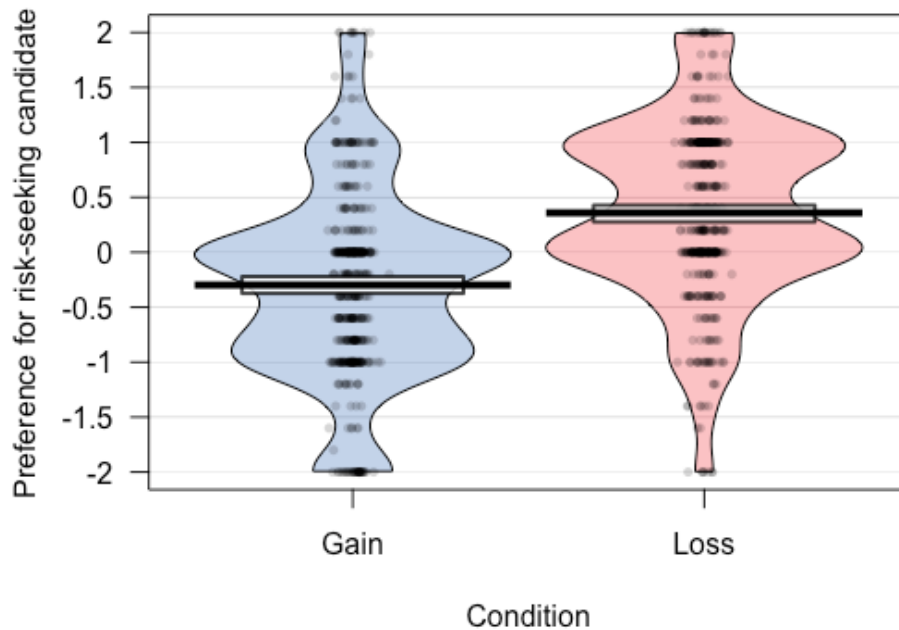
Results

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 to risk-averse policies) would be relatively more preferred when policies were framed as losses compared to 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 ($b = -0.66$, 95% CI = [-0.77, -0.55], $t = -11.66$, $p < .001$, Cohen’s $d = 0.78$). Notably, we observed a full preference reversal: while participants significantly preferred the risk-

seeking gubernatorial candidate in the loss frame ($b = 0.36$, 95% CI = [0.28, 0.44], $t = 9.25$, $p < .001$, Cohen's $d = 0.44$), this preference reversed (and was similar in magnitude) in the gain frame ($b = -0.30$, 95% CI = [-0.38, -0.22], $t = 7.32$, $p < .001$, Cohen's $d = 0.34$).

Figure 5. Preference for risk-seeking candidate as a function of frame (Experiment 3, N = 906). 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.



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 to 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 to a preference for the risk-seeking candidate just 23% of the time (the remaining 23% indicated no preference). 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 provide evidence that decision frames are socially relevant for leaders, even when the choice options are themselves equivalent across frames.

Individual difference analyses. In an exploratory analysis, we assessed whether three individual difference variables moderated the effect of 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 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: $b = 0.00$, 95% CI = [-0.05, -0.06], $t = 0.16$, $p > .87$. Second, we

found a similar null result for self-reported knowledge of framing effects ($b = -.06$, 95% CI = [-0.32, 0.19], $t = -0.48$, $p = .63$), 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 sub-set of participants who indicated they were not registered or did not plan to vote).

Discussion

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

General discussion

Framing effects on risk preferences are among the most influential findings in the social and behavioral sciences (Tversky & Kahneman, 1981; for review, see Ruggeri et al., 2020). Prior research has 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 (versus risk-seeking) choices were socially and financially penalized by third-party observers when choice outcomes were framed as losses, these same risk-averse individuals were socially rewarded when outcomes were framed as gains. 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 occur. Observers were more likely to disagree with decision makers who made frame-inconsistent choices. In turn, as predicted by theory and research on naïve realism (Pronin, Gilovich, & Ross, 2004; Ross & Ward, 1995), disagreement was associated with negative reputational consequences.

Theoretical Implications

The present research provides an important qualification to the strong prescriptive claim that decision frames should *always* be ignored. Decision makers often care deeply about how their choices are perceived by others (Lerner & Tetlock, 1999; Schlenker & Weigold, 1992; Tetlock, 2002). To the extent that individuals attempt to manage their reputation, an individual decision maker may be justified in attending to decision frames. Of note, the results hold implications not only for decision making under risk, but also for extending understanding of a range of other behavioral tendencies long considered irrational biases.

Our work also has two further preliminary theoretical implications. First, while prior work has explored cognitive drivers of framing effects (Bazerman & Moore, 2009), the reputational incentives identified in this work may illuminate a novel driver for why framing effects (and other such deviations from prescriptive decision models) occur. 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, 2019). While 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.

Second, the existence of such reputational costs suggests new potential levers to reduce framing effects – and a host of other behavioral biases. While prior research has uncovered successful strategies for de-biasing decision making by focusing on the cognition of the individual decision maker (Morewedge et al., 2015; Sellier, Scopelliti, & Morewedge, 2019), the present research suggests that an additional focus on the social environment 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. Future work is needed before any firm conclusions can be drawn about the potential effectiveness of such interventions.

Practical Implications

One key practical implication of this work is that leaders in the public and private sector could harness decision frames to increase or decrease the social palatability of candidates or policies (Zlatev, Daniels, Kim, & Neale, 2017). This may already have happened. For example, as observed in an editorial in *The New York Times* (Vavreck, 2016), in 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 use decision frames to build consensus.

Limitations and Future Directions

A few notable limitations and future directions of this work merit note. First, on a methodological level, while 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 (see Supplement Information for details). Second, the present work is limited to third-party observers. Yet, individuals who are affected by the participant's choice may have different perceptions of the decision maker. Additionally, while Experiment 2 introduced real financial stakes as a dependent variable, the decisions that observers considered were hypothetical. Observers may respond more or less strongly when they evaluate consequential choices. Finally, future research is needed to expand this work to the host of other behavioral biases in the literature on judgment and decision making (for related work, see Everett, Faber, Savulescu, & Crockett, 2018; Grossman, Eibach, Koyama, & Sahi, 2020; Tenney et al., 2019).

Conclusion

A mature literature on human judgment and decision making has established that decision frames influence risk preferences (Tversky & Kahneman, 1981; for review, see Ruggeri et al., 2020). In the present research, we provide evidence that decision frames can be reputationally relevant at the same time they are economically irrelevant. In line with research on naïve realism (Pronin, Gilovich, & Ross, 2004; Ross & Ward, 1995), results suggest this occurs because third-party observers themselves are influenced by decision frames and punish disagreement. Future research is needed to explore boundary conditions in both the lab and field, examine whether results generalize to other

behavioral biases, and test interventions based on understanding the reputational costs of economic rationality.

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.
- Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110*(3), 486.
- Bazerman, M. H. (1983). Negotiator judgment: A critical look at the rationality assumption. *American Behavioral Scientist, 27*(2), 211-228.
- Bazerman, M. H. & Moore, D. A. (2009). *Judgment in managerial decision making*. New York: Wiley.
- Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics, 110*(1), 73-92.
- Briggs, R. A. (2019). Normative Theories of Rational Choice: Expected Utility. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/archives/fall2019/entries/rationality-normative-utility/>.
- Byrne, D. (1961). Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology, 62*, 713-715.
- 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.
- Everett, J. A., 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.

- 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.
- Fischhoff, B., & Broomell, S. B. (2020). Judgment and decision making. *Annual Review of Psychology*, *71*, 331-355.
- 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.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25-42.
- Gigerenzer, G. (2014). Should patients listen to how doctors frame messages?. *BMJ*, *349*, g7091.
- Gilovich, T. D., & Griffin, D. W. (2010). Judgment and decision making. *Handbook of Social Psychology*.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.
- Grossmann, I., Eibach, R. P., Koyama, J., & Sahi, Q. B. (2020). Folk standards of sound judgment: Rationality Versus Reasonableness. *Science Advances*, *6*(2), eaaz0289.
- 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. *Journal of Risk and Insurance*, *111-132*.

- 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.
- Jordan, J. J., & Rand, D. G. (2019). 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*.
- 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.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263-291.
- Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. *Nobel prize lecture*, *8*, 351-401.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255.
- Liberman, 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.
- Loewenstein, G., Rick, S., & Cohen, J. D. (2008). Neuroeconomics. *Annual Review of Psychology*, *59*, 647-672.
- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267.
- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic.

- Psychological Bulletin*, 111, 361–365.
- McNeil, B. J., Pauker, S. G., Sox Jr, H. C., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine*, 306(21), 1259-1262.
- 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.
- 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.
- Mellers, B. A., Schwartz, A., & Cooke, A. D. (1998). Judgment and decision making. *Annual Review of Psychology*, 49(1), 447-477.
- 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.
- 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.
- 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.
- 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.
- Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. In *Advances in Experimental Social Psychology*, 27, 255-304. Academic Press.
- 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.
- Rosseel, 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.
- Ruggeri, K., Alí, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., ... & Gibson, S. P. (2020). Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, 1-12.
- Schlenker, B. R., & Weigold, M. F. (1992). Interpersonal processes involving impression regulation and management. *Annual Review of Psychology*, 43(1), 133-168.
- Sellier, A. L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychological Science*, 30(9), 1371-1379.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. Available at [SSRN 2160588](https://ssrn.com/abstract=2160588).
- Sun, Y., & Mellers, B. (2016). Trade-upgrade framing effects: Trades are losses, but upgrades are improvements. *Judgment & Decision Making*, 11(6).
- Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223(5), 96-103.
- 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.
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: intuitive politicians, theologians, and prosecutors. *Psychological Review*, *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.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458.
- Vavreck, L. (2016, May 9). Big picture for Trump vs. Clinton? It's all in the framing. *The New York Times*, Retrieved from <http://www.nytimes.com>.
- Von Neumann, J. & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
- Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, *60*, 53-85.
- 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*, *114*(52), 13643-13644.