Judgment and Decision Making, Vol. 12, No. 3, May 2017, pp. 236–252

Moral judgments of risky choices: A moral echoing effect

Mary Parkinson[*] Ruth M. J. Byrne[†]

Abstract

Two experiments examined moral judgments about a decision-maker's choices when he chose a sure-thing, 400 out of 600 people will be saved, or a risk, a two-thirds probability to save everyone and a one-thirds probability to save no-one. The results establish a moral echoing effect — a tendency to credit a decision-maker with a good outcome when the decision-maker made the typical choices of the sure-thing in a gain frame or the risk in a loss frame, and to discredit the decision-maker when there is a bad outcome and the decision-maker made the atypical choices of a risk in a gain frame or a sure-thing in a loss frame. The moral echoing effect is established in Experiment 1 (n=207) in which participants supposed the outcome would turn well or badly, and it is replicated in Experiment 2 (n=173) in which they knew it had turned out well or badly, for judgments of moral responsibility and blame or praise. The effect does not occur for judgments of cause, control, counterfactual alternatives, or emotions.

Keywords: moral echoing, risk, frame, blame, moral responsibility

1 Introduction

People often judge the morality of other people's decisions, and their tendency to do so has widespread consequences for every day social engagement, as well as for political, legal, and social policy. Our aim is to examine whether judgments of moral responsibility and blame are affected by whether a decision-maker's choice was risky or not. Our focus is on whether the decision-maker's choice was a sure-thing or a risk, whether it was framed in terms of gains or losses, and whether the outcome turned out to be good or bad. We examine the idea of "moral echoing", that is, a tendency for participants to praise a decision-maker most for a good outcome when the decision-maker made the typical choices of the sure-thing in a gain frame or the risk in a loss frame (the choices which echo those made by the typical participant), and to blame the decision-maker most when a bad outcome followed the atypical choices of a risk in a gain frame or a sure-thing in a loss frame.

1.1 Risky and sure options framed as gains and losses

We ask whether moral judgments about other people's choices echo the typical decisions that most people make. We examine whether people's moral judgments about a decision-maker's choices are affected by whether they were risky or not, which arises from the observation that people's own decisions are affected by riskiness. Consider the well-known Asian disease problem:

> Imagine that the US is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:
>
> Program A: If Program A is adopted, 200 people will be saved.
>
> Program B: If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

When the options are framed in terms of lives saved, most people chose Program A, the sure option, exhibiting a tendency to be risk averse (Tversky & Kahneman 1981). A second version of the problem provides options framed as losses:

> Program A: If Program C is adopted 400 people will die.
>
> Program B: If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

When the options are framed in terms of people dying, most people chose Program B, the risky option, exhibiting a tendency to be risk seeking for losses (Tversky & Kahneman 1981). The preference reversal is generally considered

The research was supported by a PhD scholarship from the Irish Research Council (Humanities and Social Sciences) to the first author. We thank the Science Gallery at Trinity College Dublin for their support during the "RiskLab" exhibition, and Geoff Goodwin for helpful comments. Some of the results were presented at the London Reasoning Workshop in 2014.

[*]University College Dublin, Ireland

[†]Trinity College Dublin, University of Dublin, Ireland

inconsistent because equivalent outcomes are treated differently (Kahneman & Tversky, 1979; Tversky & Kahneman 1992). The framing effect has been observed for many different contents, including moral contents such as lives saved or lost (e.g., Fagley, Coleman & Simon, 2010; Reyna, Chick, Corbin & Hsia, 2013; Spence & Pidgeon 2010; Ritov & Zamir 2014) and non-moral contents such as financial investments (Roszkowski & Snelbecker, 1990; Druckman & McDermott, 2008).

We predict that people's judgments of moral responsibility and blame or praise for a decision-maker's choice of a risk or sure thing will exhibit a framing effect, echoing their own typical decisions. Hints of moral echoing can be gleaned from the observation that people tend to be more forgiving of others who are similar to them (Burger, 1981; Chaikin & Darley, 1973), and their judgments of the morality of another person's choice are affected by their own assessment of its moral acceptability (Alicke, 2000; see also Goodwin & Darley, 2008). In moral dilemmas, people judge that the typical choices — the sure option in the gain frame and the risky option in the loss frame — are more acceptable than the atypical choices (Shenhav & Greene, 2010; see also Petrinovich & O'Neill 1996). They penalize others more for causing certain death than for imposing an equivalent risk of death (e.g. Viscusi, 2000; Sunstein, 2005). Hence we expect that people's moral judgments about a decision-maker's risky or sure choices will exhibit a moral echoing effect, modulated by their thoughts about the outcome: a decision-maker will be praised most for a good outcome after the typical choices, and will be blamed most for a bad outcome after the atypical choices.

1.2 Good and bad outcomes

We examine moral judgments about a decision-maker's choices when a future outcome is hypothesized to turn out to be good or bad in Experiment 1, and when a past outcome is known to have turned out to be good or bad in Experiment 2. We expect that decision-makers will be praised for a good outcome after typical choices — the sure-thing in a gain frame and the risk in a loss frame — and blamed for bad outcomes after atypical choices — the risk in the gain frame and the sure-thing in a loss frame. Moral responsibility is not a valenced judgment — a decision-maker can be acclaimed by being judged morally responsible (in a good way) for a good outcome and censured by being judged morally responsible (in a bad way) for a bad outcome. Hence we expect the standard framing effect for good outcomes — the decision-maker will be judged more morally responsible for a good outcome following the typical choices; and we expect a "mirror-image" framing effect for bad outcomes — the decision-maker will be judged more morally responsible for a bad outcome when the decision-maker makes the atypical choices.

Knowledge that an outcome was good or bad affects judgments of the quality of a decision-maker's thinking, and his or her competence, even though people know they should not consider outcomes in such evaluations (Baron & Hershey, 1988). This outcome bias may arise because a bad outcome tends to overshadow an assessment of the decision-maker's intentions (Sezer, Zhang, Gino & Bazerman 2016). The good or bad outcome may call attention to arguments that make the choice seem good or bad (Baron & Hershey 1988). A similar mechanism has been proposed for hindsight bias, the tendency for people with outcome knowledge to believe they would have predicted the outcome (Fischhoff, 1975; Hawkins & Hastie 1990; Roese & Vohs 2012). The outcome of a decision can affect people's judgments of its morality. An agent's decision is considered more important for a good outcome than a bad outcome, and people are generally assumed to intend good outcomes (Pizarro, Uhlman & Salovey, 2003). The worse the outcome is, the greater the tendency for people to judge an agent to be responsible and blameworthy (e.g. Walster, 1966; Burger, 1981; Darley & Shultz, 1990; see also Mitchell & Kalb, 1981; Ames & Fiske, 2013). Hence, we expect that moral judgments of responsibility and praise or blame will exhibit a moral echoing effect.

1.3 Moral judgments and non-moral judgments

We examine the idea of moral echoing for judgments of moral responsibility and blame or praise for a decision-maker. We also examine non-moral judgments about whether the decision-maker caused the outcome and was in control of it, judgments about how it could have turned out differently, and judgments about whether the decision-maker experienced relief or upset about the outcome. We suggest that moral echoing may be confined to inferences about moral and social regulations, and it may arise because people infer that the moral principles they apply to their own decisions are the same principles that others should apply to theirs (Haidt 2001; Mikhail 2007). Hence, we do not expect to observe an echoing effect for judgments outside the moral or social realm: we do not expect that people will judge that a decision-maker causes or controls a good outcome when the decision-maker makes the typical choices of the sure-thing in a gain frame or the risk in a loss frame, and causes or controls a bad outcome following the atypical choices of a risk in a gain frame or a sure-thing in a loss frame. Their moral judgments are based on deontic inferences about what should happen, which may be particularly sensitive to perspective (Quelhas & Byrne 2003; Rasga, Quelhas & Byrne 2016), whereas their judgments of causality or controllability are based on epistemic inferences about what can happen.

2 Experiment 1

The aim of the experiment was to examine whether participants' moral judgments about a decision-maker's choices are affected by (a) whether the decision-maker choses a sure option or a risky option, (b) whether the options are framed as gains or losses, and (c) whether the participant supposes that the outcome is good or bad. We focused on participants' moral judgments of blame or praise and their judgments of moral responsibility. We also examine non-moral judgments of cause, control, counterfactual thoughts, and emotions.

2.1 Method

2.1.1 Participants

The participants were 207 volunteers recruited via the online platform, Crowdflower at Crowdflower.com.[1] There were 63 women and 144 men, aged between 18 to 59 years. A further 3 participants were eliminated prior to any data analysis on the basis of three criteria used to screen participants: (1) participants were asked to select the name of the protagonist from a list of four names, and asked how many lives were at stake in total; participants who answered both questions incorrectly were eliminated, (2) participants were asked to participate only if they had not taken part in something similar before; they were asked at the end of the experiment whether they had done something similar before and if so, what its subject was, and participants who had taken part in something similar before were eliminated, (3) participants with identical worker IDs were eliminated. Participants were assigned at random to one of eight conditions, which varied the decision-maker's choice, the frame, and the outcome: sure gain bad-outcome n=25 and risky gain bad-outcome n=21, sure loss bad-outcome n=23 and risky loss bad-outcome n=25, sure gain good-outcome n=43 and risky gain good-outcome n=21, sure loss good-outcome n=23 and risky loss good-outcome n=26. Participants were given a nominal $.15 for their participation.

2.1.2 Materials and design

All participants were presented with the Asian disease problem. The design was a fully between-participants one: 2 (decision-maker's choice: sure-thing vs. risk) x 2 (frame: gain vs. loss) x 2 (hypothetical outcome: good vs. bad). Participants in the sure-thing conditions were informed that the decision-maker, John, had chosen program A (the sure option), and those in the risky conditions were told that John had chosen program B (the risky option). For participants in the gain conditions the options were framed in terms of gains, that is, people being saved, and for those in the loss conditions the equivalent options were framed in terms of losses, that is, people dying. Participants in the good outcome conditions were implicitly asked, through the phrasing of question, to suppose things turned out well: "John would be morally responsible for people being saved." Those in the bad outcome conditions were asked to suppose things turned out badly: "John would be morally responsible for people dying." In addition, the judgments about the hypothetical outcome were framed in terms of gains and losses appropriately for each condition, i.e., in the gain conditions the judgments referred to people being saved (good outcome) or not being saved (bad outcome) whereas in the loss conditions the judgments referred to people dying (bad outcome), or not dying (good outcome). Appendix A gives the complete text.

Before they were given any information about which option the decision-maker had chosen, participants first rated each of the two options as morally acceptable on a likert scale anchored at 1=completely disagree, and 5=completely agree to check that participants tended to make the typical choices of a sure-thing in a gain frame, and a risk in a loss frame. Participants then made 6 judgments about the decision-maker's choice, on 1–5 scales anchored at 1=completely disagree and 5=completely agree (illustrated here for the good outcome gain frame). Two of the judgments were moral judgments: (1) John would be morally responsible for people being saved, and (2) John would deserve to be praised for people being saved. The remaining four were non-moral judgments, (3) John would cause people to be saved, (4) John would be in control of people being saved, (5) John would be relieved about people being saved. The wording of the judgments was identical in all conditions except that the final words were framed according to the condition, e.g., "... for people being saved/not being saved" in the gain frame for good and bad outcomes respectively and "... for people dying/not dying" in the loss frame for bad and good outcomes respectively; in the good outcome conditions the blame/praise judgment referred to praise, and the relief/upset judgment referred to relief, and in the bad outcome conditions the judgments referred to blame and upset respectively. The sixth measure was a composite counterfactual score derived from 4 judgments about imagined alternatives. They were instructed, "Imagine you are a member of a committee formed to review the preparations for the outbreak and to predict how things could turn out. What is the most likely way you would complete the thought, "Things could turn out differently if... ". They provided their judgments on a 1–5 scale (1=completely disagree and 5=completely agree) for each of 4 counterfactuals: (a) "... if John took more risks", (b) "... if John were a more moral person", (c) "... if John recommends the other program", and (d) "... if there were additional programs available". They received the judgments

[1] The experiment was first carried out with 253 participants, 135 volunteers recruited from the general public attending an exhibition on risk at Trinity College's Science Gallery and 118 students from the campus of Trinity College Dublin. However because the participants were assigned in two phases to the eight conditions of the experiment rather than at random, the experiment was re-run on the advice of the reviewers.

in the fixed order of responsibility, relief/upset, blame/praise, cause, control, and counterfactuals.

2.1.3 Procedure

The materials were presented via SurveyGizmo (see http://www.surveygizmo.com/, which participants were directed to from the online platform Crowdflower https://www.crowdflower.com/). The participants were asked to read the story carefully and to answer the questions in the order in which they were asked. They were told there were no right or wrong answers. The scenario was presented on screen and after participants had read it they pressed the "next" button on screen to read each of the judgment tasks. Each judgment task was presented separately on screen and participants provided their answer by clicking on one of the scale numbers on screen. The scenario remained on screen throughout. The experiment took approximately 10 minutes to complete.

2.2 Results

2.2.1 Moral acceptability

Participants were required at the outset to first indicate their own judgment of the moral acceptability of the risk and sure-thing options, as a check that they tended to make the typical choices, before they read about the decision-maker's choice. Their judgments showed a framing effect, as indicated in a 2 (frame: gain vs. loss) x 2 (choice: sure vs. risky option) ANOVA with repeated measures on the second factor. There was an interaction of choice and frame F(1,199)=24.49, p<.001, η_p^2=.11, as well as a main effect of each one: choice, F(1,199)=4.11, p<.05, η_p^2=.02, frame, F(1,199)=19.37, p<.001, η_p^2=.09. Contrasts, with a Bonferroni corrected alpha of p<.02 for 4 comparisons, showed the interaction arises because participants judged the risk to be more morally acceptable than the sure-thing when the outcomes were framed as losses, F=(1,96)=32.31, p<.001, η_p^2=.25, the sure option to be more morally acceptable when it was framed as a gain than a loss F(1,205)=47.06, p<.001, η_p^2=.19, and no other comparisons were significant (largest F=3.62, smallest p<.06), as Figure 1A shows.

For each of the 6 judgments about the decision-maker we report first the effects of outcome in a 2 (hypothetical outcome: good vs. bad) x 2 (decision-maker's choice: sure vs. risky) x 2 (frame: gain vs. loss) between participants ANOVA. Next, we test framing effects for good outcomes and for bad outcomes in a 2 (decision-maker's choice: sure vs. risky) x 2 (frame: gain vs. loss) ANOVA. We also report correlations of each judgment with the other judgments.

Results on the relationship between participants' own judgments and their answers to other questions are described in Appendix B, for both experiments. These results do not affect any conclusions described in the main text.

Table 1: Table 1: Correlations between the judgments about a decision-maker's choice in Experiments 1 and 2.

	Blame	Cause	Control	Counter-factuals	Upset
Experiment 1, n=207, r=.12 for p<.05 one tailed					
Responsibility	.41	.33	.19	.22	.03
Blame		.66	.37	.08	.02
Cause			.42	.08	.05
Control				.20	.06
Counterfactuals					.01
Experiment 2, n=173, r=.13 for p<.05 one tailed					
Responsibility	.57	.51	.53	.20	.13
Blame		.54	.51	.10	.14
Cause			.45	.19	.18
Control				.14	–.07
Counterfactuals					.00

2.2.2 Moral judgments

A moral echoing effect was found for moral responsibility and blame and praise judgments, that is, participants tended to credit the decision-maker most for a good outcome when he made the typical choices of the sure-thing in a gain frame or the risk in a loss frame, and to discredit him most when there was a bad outcome and he made the atypical choices of a risk in a gain frame or a sure-thing in a loss frame.

Moral responsibility. There was a main effect of outcome, F(1,199)=12.03, p <.001, η_p^2=.06, as the decision-maker was judged more responsible for good outcomes than bad ones, and outcome, choice, and frame interacted F(1, 199)=9.02, p< .01, η_p^2=.04; there were no other differences (largest F=.3, smallest p<.58). Contrasts, with a Bonferroni corrected alpha of p<.004 for 12 comparisons, showed the decision-maker was judged morally responsible for a good outcome more than a bad outcome when he chose the risk in a loss frame, F(1,199)=12.28, p<.001, η_p^2=.058, or the sure-thing in a gain frame, F(1,199)=11.69, p<.001, η_p^2=.055. No other contrasts from the set of 12 showed significant differences (largest F=3.65, smallest p<.058). Next we assessed framing effects: moral responsibility judgments showed a standard framing effect for good outcomes, the decision-maker was judged more morally responsible for a good outcome when he made the typical choices, a sure-thing in a gain frame, a risk in a loss frame, as the interaction of frame and choice shows, F(1,109)=4.38, p<.05, η_p^2=.039; and they showed a mirror-image framing effect for bad outcomes, the decision-maker was judged more morally responsible for a bad outcome when he made atypical choices, the risk in a gain frame

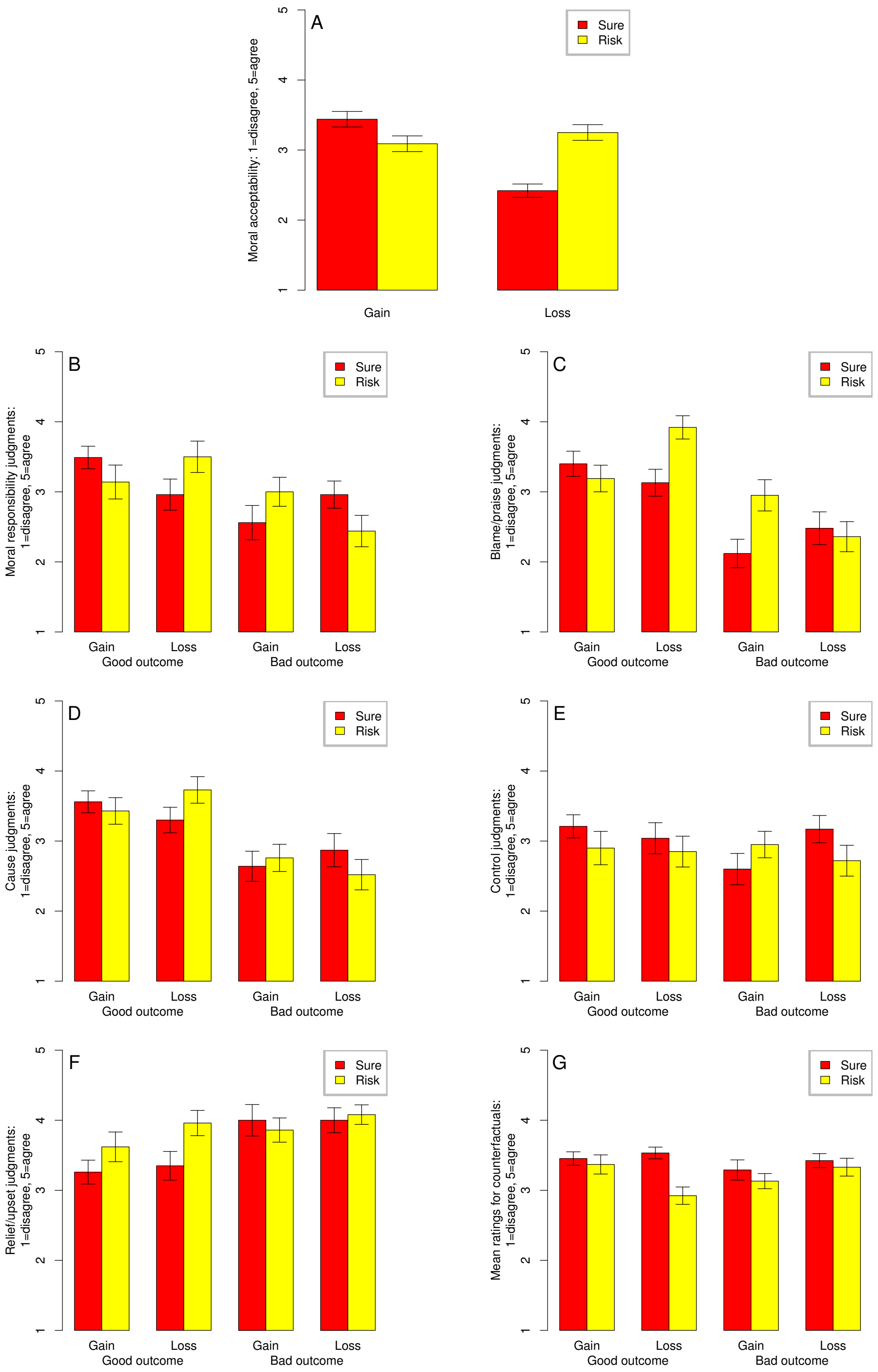

Figure 1: Judgments of (a) initial acceptability, (b) moral responsibility, (c) blame/praise, (d) cause, (e) control, (f) relief/upset, and (g) counterfactuals for sure and risky options in gain and loss frames for supposed good and bad outcomes in Experiment 1. Error bars are standard error of the mean.

or the sure-thing in a loss frame, as the interaction of frame and choice shows, F(1,90)=4.66, p<.05, η_p^2=.049, as Figure 1B illustrates. Moral responsibility judgments correlated with blame/praise judgments, r=.41, p<.001; cause r=.33, p<.001, control r=.19, p<.01, and counterfactual judgments r=.22, p<.01, but not with relief/upset judgments r=.03, p<.63, as Table 1 shows.

Blame/praise. There was a main effect of outcome F(1,199)=40.47, p<.001, η_p^2=.17, as the decision-maker was praised for good outcomes more than he was blamed for bad ones, and of choice F(1,199)=4.93, p<.05, η_p^2=.02, as the decision-maker was judged more deserving of blame/praise when he chose the risky option than the sure one, and outcome, choice, and frame interacted F(1,199)=11.05, p<.001, η_p^2=.05; there were no other differences (largest F=1.43 smallest p<.23). Contrasts show the decision-maker was praised for a good outcome more than blamed for a bad one when he chose the risk in a loss frame F(1, 199)=29.36, p<.001, η_p^2=.13, or when he chose the sure thing in a gain frame, F(1,199)=24.24, p<.001, η_p^2= .11. He was praised most for a good outcome when he chose the risk rather than the sure-thing in a loss frame F(1,199)=7.23, p<.008, η_p^2=.035, he was blamed least for a bad outcome when he chose the sure-thing rather than the risk in a gain frame F(1,199)=7.46, p<.007, η_p^2=.036. No other contrasts from the set of 12 showed significant differences on the Bonferroni corrected alpha (largest F=5.88, smallest p=.016). Next, the praise judgments show a standard framing effect for good outcomes, the decision-maker was praised for a good outcome when he made the typical choices, a sure-thing in a gain frame, a risk in a loss frame, as the interaction of frame and choice shows, F(1, 109)=6.45, p<.05, η_p^2=.056; and the blame judgments showed a mirror-image framing effect for bad outcomes, the decision-maker was blamed more for a bad outcome when he made atypical choices, the risk in a gain frame or the sure-thing in a loss frame, as the interaction of frame and choice shows, F(1,90)=4.66, p<.05, η_p^2=.049, as Figure 1C illustrates. Blame/praise judgments correlated with moral responsibility judgments as reported above, and with cause r=.66, p<.001 and control r=.37, p<.001, but not with counterfactual judgments r=.08, p<.23, or relief/upset judgments r=.02, p<.8.

The results for moral judgments show a moral echoing effect: a decision-maker is credited with a good outcome when he made the typical choices, he is discredited when there is a bad outcome and he made atypical choices, the risk in a gain frame or the sure-thing in a loss frame.

2.2.3 Non-moral judgments

There is no echoing effect for the non-moral judgments of cause, control, counterfactuality and relief or upset. They show a strong effect of outcome, but for the most part, no framing effects.

Causal judgments. There was a main effect of outcome F(1, 199)=32.03, p<.001, η_p^2=.14, as the decision-maker was judged to have caused good outcomes more than bad ones; and no other differences (largest F=3.24, smallest p<.072). Causal judgments did not show a framing effect, there was no interaction of frame and choice for good outcomes, F (1,109)=2.21, p=.14, and none for bad outcomes, F(1,90)=1.16, p=.28, as Figure 1D illustrates. Causal judgments correlated with moral responsibility and blame/praise judgments as reported above, and with control judgments r=.42, p<.001, and with counterfactual judgments r=.18, p<.01 but not with relief/upset judgments r=.05, p<.49.

Control judgments. There were no effects of outcome, choice, or frame (largest F=2.3, smallest p<.13). Control judgments did not show a framing effect, there was no interaction of frame and choice for good outcomes, F<1, and none for bad outcomes F(1,90)=3.67, p=.059, η_p^2=.04, as Figure 1E shows. Control judgments correlated with moral responsibility, blame/praise judgments and cause as reported above, and with counterfactual judgments r=.20, p<.01; they did not correlate with relief/upset judgments r=.06, p<.42.

Relief/upset judgments. There was a main effect of outcome, F(1,199)=10.28, p<.01, η_p^2=.05, as the decision-maker was judged to be more upset about a bad outcome than relieved about a good outcome, and there were no other effects (largest F=3.62, smallest p=.06). Relief/upset judgments did not show a framing effect, there was no interaction of frame and choice for good outcomes, F <1, and none for bad outcomes, F<1, as Figure 1F shows. Relief/upset judgments did not correlate with any of the other judgments, as reported above, nor with counterfactual judgments r=.014, p<.83.

2.2.4 Counterfactuals

The composite score of the overall agreement with the four counterfactuals showed no main effect of outcome, a main effect of choice, F(1,199)=7.85, p<.01, η_p^2=.04, outcome interacted with frame, F(1, 199)=4.29, p<.05, η_p^2=.02; and there were no other differences (largest F=3.05, smallest p < .08). The counterfactual judgments show a framing effect when participants supposed the outcome was good, as the interaction of frame and choice for good outcomes shows, F(1,109)=5.18, p<.05, η_p^2=.05; there was no interaction of frame and choice for bad outcomes, F<1. Counterfactual judgments correlated with moral responsibility and control judgments as reported above, but not with blame/praise, cause, or relief/upset judgments also as reported above.

2.3 Discussion

Decision-makers were not judged to have caused or to be in control, or to feel relief for a good outcome when they chose the sure-thing in a gain frame, and the risk in a loss frame, nor were they judged to have caused or to be in control of, or to feel upset for, a bad outcome when they chose the risk in a gain frame, and the sure-thing in a loss frame. Decision-makers were judged to have alternatives when the outcome was good and they chose the sure-thing in a gain frame, and the risk in a loss frame, but not when the outcome was bad.

The results show that participants own moral judgments, before they hear about what the decision-maker did, show a framing effect. Their moral judgments about a decision-maker who makes these choices also show a framing effect, which depends on their supposition about whether the outcome is good or bad. This moral echoing effect leads them to credit a decision-maker for a good outcome when he chose the sure-thing in a gain frame, and the risk in a loss frame, and to discredit him when there is a bad outcome and he chose the risk in a gain frame, and the sure-thing in a loss frame.

It is noteworthy that moral and non-moral judgments show a strong effect of outcome. Participants judged the decision-maker to be more morally responsible for good outcomes than bad ones, and they praised him more for good outcomes than they blamed him for bad ones. Their non-moral judgments also showed an effect of outcome, they judged him to cause good outcomes more than bad ones, and they considered he would be upset about a bad outcome more than relieved about a good outcome; there were no effects of outcome for judgments of control or for counterfactual alternatives. The next experiment examines the moral echoing effect for *known* good and bad outcomes.

3 Experiment 2

The aim of the experiment was to replicate and extend the moral echoing effect discovered in Experiment 1 for moral judgments about the decision-maker's choices when participants were told that the outcome was good or bad.

3.1 Method

3.1.1 Participants

The participants were 173 volunteers recruited from the online platform, Crowdflower (see Crowdflower.com).[2] There were 53 women and 120 men, aged between 18 and 75 years. A further 6 participants were eliminated prior to any data analysis on the basis of the same 3 criteria used to screen participants in the previous experiment. Participants were assigned at random to one of eight conditions: sure gain good-outcome n=21 and risky gain good-outcome n=20, sure loss good-outcome n=23 and risky loss good-outcome n=23, sure loss bad-outcome n=24 and risky loss bad-outcome n=20, sure gain bad-outcome n=22 and risky gain bad-outcome n=20. They were compensated $.15 for their time.

3.1.2 Materials, design and procedure

The materials, design and procedure were the same as the previous experiment except that participants were told the outcome had turned out well, framed as "As a result of his decision a lot of people were saved" (gain frame), or "As a result of his decision a lot of people did not die" (loss frame), or they were told it had turned out badly, framed as "As a result of his decision a lot of people were not saved" (gain frame), or "As a result of his decision a lot of people died" (loss frame). Participants made the same judgments as the previous experiment. The imagined alternatives in the experiment were counterfactuals about the past, i.e., they were phrased "Things could have turned out differently if. . . ".

3.2 Results

3.2.1 Moral acceptability

As in the previous experiment participants tended to make the typical choices: there was a main effect of frame, F(1,171)=3.99, p<.05, η_p^2=.02, no main effect of choice F(1,171) =3.54, p<.062, and an interaction of the two F(1,171)=30.98, p<.001, η_p^2=.15, as Figure 2A shows. Once again contrasts show that participants judged the sure thing to be more morally acceptable than the risk when the choices were framed as gains, F(1,82) =13.41, p<.001, η_p^2=.14 and the risk to be more acceptable than the sure thing when the choices were framed as losses, F(1,89) =17.87, p<.001, η_p^2=.17. The sure thing was more acceptable when it was framed in terms of gains than losses F(1,169)=30.44, p<.001, η_p^2=.15, and the risk was more acceptable when it was framed in terms of losses than in terms of gains F(1,169)=9.07, p<.01, ηp^2 =.05. The same set of analyses were carried out as in the previous experiment on the 6 judgments about the decision-maker.

A moral echoing effect was again found for moral responsibility and blame and praise judgments, replicating the results of the first experiment and extending them to known outcomes.

Moral responsibility. As in Experiment 1, there was a main effect of outcome, F(1,165)=14.8, p<.001, η_p^2=.08, an interaction of frame and outcome F(1,165)= 7.96, p<.001,

[2] The experiment was first carried out with 188 volunteers recruited from the general public attending an exhibition on risk at Trinity College's Science Gallery. However, because it contained only four of the eight possible conditions, the experiment was re-run on the advice of the referees.

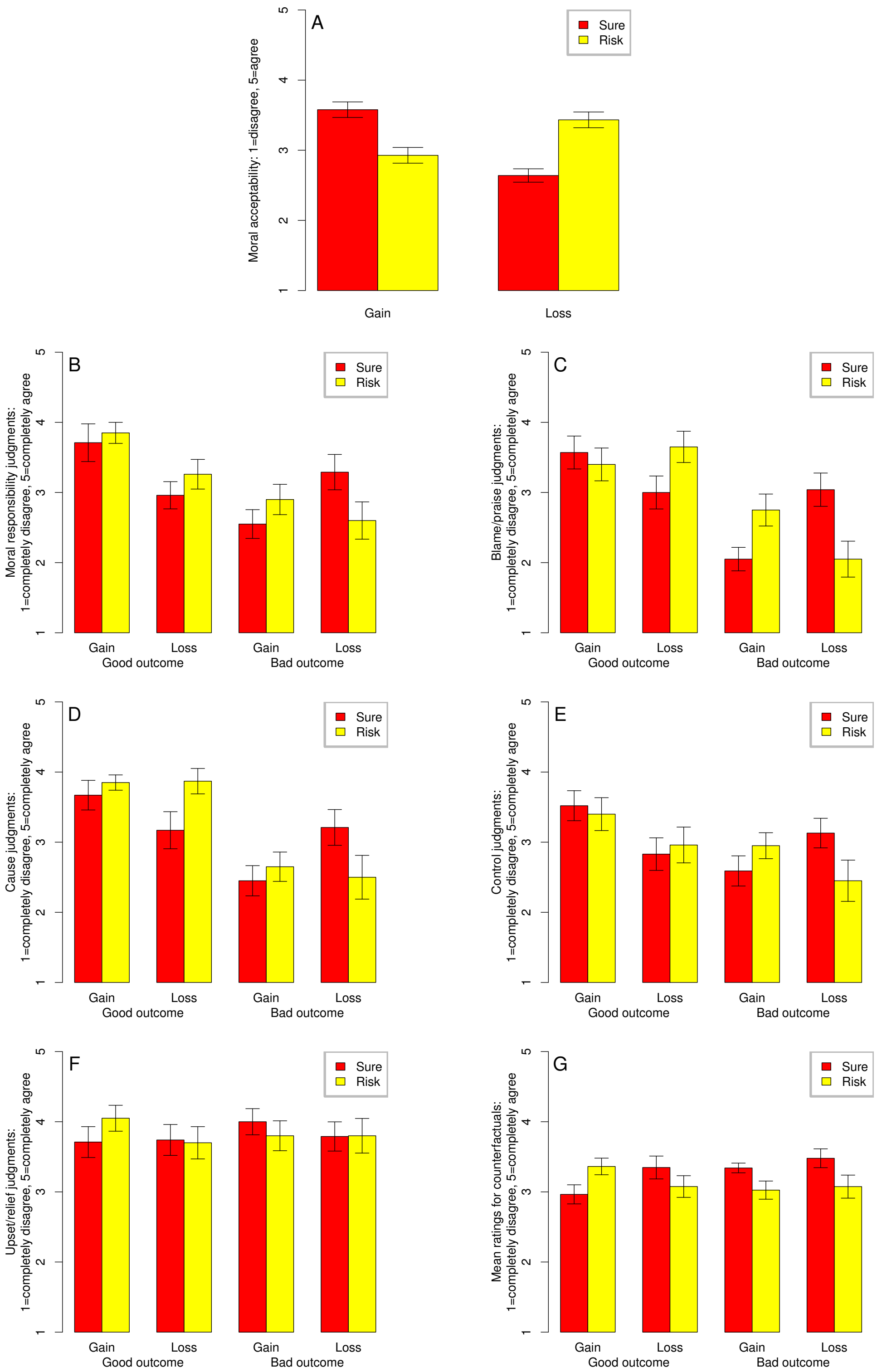

Figure 2: Judgments of (a) initial moral acceptability, (b) moral responsibility, (c) blame/praise, (d) cause, (e) control, (f) relief/upset, and (g) counterfactuals for sure and risky options in gain and loss frames for known good and bad outcomes in Experiment 2. Error bars are standard error of the mean.

η_p^2=.046, a marginal interaction of frame, choice and outcome F(1,165) = 3.66, p<.058, η_p^2=.022, and no other differences (largest F=2.01, smallest p<.16). The three way marginal interaction arises because the decision-maker was judged responsible for good outcomes more than bad ones for the sure thing in a gain frame F(1,165)=13.51, p<.001, η_p^2=.08, and for the risk in a loss frame F(1, 165)=4.3, p=.04, η_p^2 =.025, and for the risk in a gain frame F(1,165)=8.31, p<.004, η_p^2=.048. No other contrasts showed significant differences on the Bonferroni corrected alpha for 12 comparisons (largest F=5.88 smallest p<.016). Next, moral responsibility judgments did not show a framing effect for good outcomes, there was no interaction of frame and choice, F<1, but there was a mirror-image framing effect for bad outcomes, the decision-maker was judged more morally responsible for a bad outcome following an atypical choice, the risk in a gain frame or the sure-thing in a loss frame, as the interaction shows, F(1,82) =4.84, p<.05, η_p^2=.056 (Figure 2B). Moral responsibility judgments correlated with blame/praise judgments r=.57, p<.001, and with cause r=.51, p<.001, control r=.53, p<.001, and counterfactual judgments r=.2, p<.01, but not with relief/upset judgments, r=.13, p<.09, as Table 1 shows.

Blame/praise. As in Experiment 1, there was a main effect of outcome, F(1,165)=33.36, p<.001, η_p^2=.17, an interaction of frame, choice and outcome F(1,165)=15.17, p<.001, η_p^2=.08, and no other differences (largest F =1.41, smallest p<.24). The interaction arises because the decision-maker was blamed more for a bad outcome when he chose the sure-thing than the risk in a loss frame F(1,165) =9.53, p<.002, η_p^2=.055, and more in the loss frame than a gain frame F(1,165)=10.12, p<.01, η_p^2=.058; he was praised for a good outcome more than blamed for a bad outcome when he chose the risk in a loss frame, F(1,165) =24.5, p<.001, η_p^2=.13, and the sure-thing in a gain frame, F(1,165)=22.22, p<.001, η_p^2=.12. No other contrasts showed significant differences on the Bonferroni corrected alpha for 12 comparisons (largest F=4.62, smallest p<.033). Next, praise judgments showed a standard framing effect for good outcomes, the decision-maker was praised for a good outcome when he made the typical choices, a sure-thing in a gain frame, a risk in a loss frame, as the interaction of frame and choice for good outcomes shows, F(1,83) =3.14, p<.08, η_p^2=.036; and the blame judgments showed a mirror-image framing effect for bad outcomes, the decision-maker was blamed more for a bad outcome following an atypical choice, the risk in a gain frame or the sure-thing in a loss frame, as the interaction of frame and choice shows F(1,82) =14.23, p<.001, η_p^2=.15, see Figure 2C. Blame/praise judgments correlated with moral responsibility judgments, as described above, and with cause r=.54, p<.001, and control r=.51, p<001, but not with relief/upset judgments r=.14, p<.08 or counterfactual judgments r=.10, p<.21.

There was no echoing effect for the non-moral judgments of cause, control, counterfactuality and relief or upset, as Figures 2D–G show. The non-moral judgments show a strong effect of outcome, but for the most part, no framing effects, replicating Experiment 1. For brevity we report these results in Appendix C.

3.3 Discussion

The results replicate the findings of the first experiment for supposed good and bad outcomes and extend them to known good and bad outcomes. Participants' moral judgments once again show a moral echoing effect that leads them to credit a decision-maker for a good outcome when the decision-maker chose the sure-thing in a gain frame or the risk in a loss frame, and to discredit the decision-maker when a bad outcome followed the choice of the risk in a gain frame or the sure-thing in a loss frame. Once again, it is noteworthy that participants judged the decision-maker to be more morally responsible for known good outcomes than known bad ones and to be more worthy of praise for good outcomes than blame for bad ones.

4 General Discussion

The two experiments show that people's judgments of the morality of another person's decisions are affected by whether the decision-maker chose a sure-thing or a risk, framed in terms of losses or gains. Participants' own moral judgments, before they hear about what a decision-maker did, show a framing effect, they judge that a risk is more morally acceptable than a sure-thing when the outcomes are framed as prospective losses, lives that may be lost, and a sure-thing is more morally acceptable than a risk when the outcomes are framed as prospective gains, lives that may be saved, as both experiments show. A striking result is that their judgments about a decision-maker's choices show a moral echoing effect, that is, a tendency to credit a decision-maker most for a good outcome when the decision-maker made the typical choices of the sure-thing in a gain frame or the risk in a loss frame, and to discredit the decision-maker most when there is a bad outcome and the decision-maker made the atypical choices of a risk in a gain frame or a sure-thing in a loss frame. Hence, participants' moral judgments about a decision-maker who makes the typical choices show a framing effect that differs depending on whether the outcome is good or bad. When they suppose or know the outcome turned out well, their judgments show the standard framing effect: they praise the decision-maker and judge the decision-maker to be morally responsible for making the typical choices; when they suppose or know the outcome turned out badly, their judgments show a mirror-image framing effect: they blame the decision-maker and judge the decision-maker morally responsible for making the atypical choices. There is no echoing effect for non-moral judgments about whether the decision-maker caused the outcome, controlled it, felt relieved or upset about it, or for counterfactuals about how it could have turned out differently.

Another striking result is that judgments about the morality of the decision-maker's choice were affected by whether the outcome was good or bad. Participants judged the decision-maker to be more morally responsible for good outcomes than bad ones, they praised the decision-maker more for good outcomes than they blamed the decision-maker for bad ones, they judged the decision-maker to cause good outcomes more than bad ones, and they considered that a decision-maker would be upset about a bad outcome more than relieved about a good outcome. There were few effects of good or bad outcomes for judgments about whether the decision-maker controlled the outcome and for counterfactuals about how it might have turned out differently.

The moral echoing effect may reflect an essentially rational assessment by participants that the risky or sure option is "right" or "wrong" depending on the gain or loss frame. They tend to prefer the sure option in a gain frame, and the risky option in a loss frame, and so they may simply consider the risky option to be "right" in a loss frame and "wrong" in a gain frame, and they may consider the sure option to be "right" in a gain frame and "wrong" in a loss frame. The moral echoing effect may occur because participants judge favorably a decision-maker who makes the "right" choice, that is, the option that the participant evaluates as the best available — the risky option in the loss frame, or the sure option in the gain frame. Similarly, they judge unfavorably a decision-maker who makes the "wrong" choice — the risky option in the gain frame, or the sure option in the loss frame. Hence, a fruitful avenue for future studies could be to examine directly the relation between a decision-maker's own choice and their judgment of another person's choice in a large-scale correlational analysis, as well to test the moral echoing effect for moral contents other than harm such as the domain of purity violations (e.g., Parkinson & Byrne, 2017). We consider three alternative accounts of the cognitive processes that may result in the moral echoing effect next.

4.1 The mental representation of possibilities

One possible account is that the moral echoing effect occurs because of the cognitive processes that construct models of the risky and sure options (McCloy, Byrne & Johnson-Laird 2010). The framing effect itself may arise because, when people make judgments about risky and sure options, they mentally represent the possible choices in different ways. Risky options require people to consider multiple alternatives whereas sure options require them to consider a single possibility. The sure option, "If Program A is adopted, 200 people will be saved" can be mentally represented by envisaging a single possibility, whereas the risky option "If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved" is mentally represented by envisaging several alternative possibilities, and the possibilities may be annotated to capture the likelihood of each one (e.g., 1/3, or 2/3), as Table 2 (1), shows (McCloy, et al., 2010; Johnson-Laird, Legrenzi, Girotto & Legrenzi, 1999; Johnson-Laird, Khemlani & Goodwin, 2015).

Framing the options as gains or losses affects which information is explicitly included in the mental representation of the different possibilities. When the options are presented as gains, the sure and risky options are mentally represented by envisaging partial information, that is, only the information corresponding to gains (lives saved). In this partial representation, the sure option explicitly represents only a good possibility (200 saved), whereas the risky option represents one good possibility (600 saved) and one bad possibility (0 saved), as Table 2(2) shows. The representation of partial information about the sure-thing and the risk makes some aspects of each option more salient than others, e.g., it makes salient the good aspects of the sure-thing. In contrast, when the options are presented as losses, the sure and risky options are mentally represented by envisaging different partial information, that is, only the information corresponding to losses (lives lost). In this partial representation, the sure option explicitly represents only a bad possibility (400 lost), whereas the risky option represents one bad possibility (600 lost) and one good possibility (0 lost). Once again the representation of partial information about the sure-thing and the risk makes some aspects of each option more salient than others, e.g., it makes salient the bad aspects of the sure-thing, as Table 2 shows.

Thus, the cognitive mechanism that leads to the observed preference reversal may arise from the representation of partial information for gains and losses. People choose the sure option when the options are framed as gains because it contains only an explicit representation of a good possibility (200 saved) whereas the risky option contains an explicit representation of one bad possibility (0 saved); in contrast they choose the risky option when the options are framed as losses because the sure option contains only an explicit representation of a bad possibility (400 lost), whereas the risky option contains an explicit representation of one good possibility (0 lost). On this account, the process of representing information parsimoniously leads to the framing effect, which may be overcome by the process of deliberately fleshing out the full information to represent it explicitly.

Finally, the moral echoing effect may arise because, when people suppose or know the outcome is good or bad, they may keep track of the epistemic status of the possibilities as facts or as counterfactual alternatives corresponding to

Table 2: Possibilities envisaged for risky and sure options framed as gains or losses.

People represent a single possibility for the sure option; and multiple possibilities for the risky option:					
	Sure option:	200 saved	400 lost		
	Risky option:	600 saved	0 lost	1/3	
		0 saved	600 lost	2/3	
People represent different information explicitly for gains and losses, e.g.:					
Gains:	Sure option:	200 saved			
	Risky option:	600 saved		1/3	
		0 saved		2/3	
Losses:	Sure option:		400 lost		
	Risky option:		0 lost	1/3	
			600 lost	2/3	
People keep track of possibilities as facts or counterfactual alternatives for good or bad outcomes:					
	Sure option:	200 saved	400 lost		FACTS
	Risky option:	600 saved	0 lost	1/3	Counterfactual –better
		0 saved	600 lost	2/3	Counterfactual –worse
	Sure option:	200 saved	400 lost		Counterfactual –worse
	Risky option:	600 saved	0 lost	1/3	FACTS
		0 saved	600 lost	2/3	Counterfactual –worse
	Sure option:	200 saved	400 lost		Counterfactual –better
	Risky option:	600 saved	0 lost	1/3	Counterfactual –better
		0 saved	600 lost	2/3	FACTS

better or worse alternatives. For example, suppose the outcome was good. When you are told that the risky option was chosen, it implies all 600 lives were saved (and so the facts are that 600 people were saved, the other possibilities are now counterfactual possibilities). The counterfactual possibilities are that things could only have turned out worse — if the risky option had not turned out well, or if the sure option had been chosen, as Table 2 step 3 shows. In contrast, when you are told that the sure option was chosen, it implies that 200 lives were saved (and so the facts are that 200 people were saved and the other possibilities are now counterfactual). The counterfactual possibilities are that things could have turned out better (if the risky option had turned out well), or they could have turned out worse (if the risky option had not turned out well). Hence, when you suppose a good outcome, when the risky option was chosen, things could only have turned out worse; when the sure option was chosen things could have turned out better or worse, as Table 2(3), shows. Hence, thinking about counterfactual possibilities when the outcome was good may lead the risky option to be perceived to have been the best one and a decision-maker will be praised most and considered most morally responsible for choosing it.

A different conclusion follows when the outcome was bad. When you are told that the risky option was chosen, that implies that no lives were saved (0 saved is the facts, the other options are counterfactual). The counterfactual possibilities are that things could only have turned out better, if the risky option had turned out well, or if the sure option had been chosen. But when you are told the sure option was chosen that implies that 200 lives were saved, and the counterfactual possibilities are that things could have turned out better — if the risky option had turned out well — or they could have turned out worse — if the risky option had not turned out well — as outlined earlier. Hence, thinking about counterfactual possibilities when the outcome is known to have been bad may lead people to perceive the sure option to have been the best one and a decision-maker to be blamed least and considered least morally responsible for choosing it.

4.2 Biases

Another possibility is that the differential effects of good and bad outcomes arise because of biases in judgment such as an outcome bias (Baron & Hershey, 1988; Alicke, 1992, 2000). A simple outcome bias account should predict that judgments of cause and control would show an amplified judgment pattern for risky choices, compared to sure ones because the risky choice results in either the best outcome (most number of lives saved) or the worst outcome (least number of lives saved). However, in both experiments reported here, participants judged agents to be more causal and in control when they chose the sure option than the risky one, for both good and bad outcomes. The data go against the predictions of an outcome bias account. The results suggest that people do not judge that the decision-maker is more responsible, or deserving of praise or blame for the risky option, also contrary to an outcome bias account. Instead, moral judgments appeared to favor the most "acceptable" choices dependent on framing, for example, blaming the decision-maker more for a bad outcome in a loss frame if he chose the sure option than the risky option, even though the risky option resulted in a greater number of lives lost.

The effects of outcomes may also reflect a sort of hindsight bias, the tendency for people with outcome knowledge to believe they would have predicted the outcome (Fischhoff, 1975; Hawkins & Hastie, 1990; Roese & Vohs, 2012). However, there were few differences between the supposed outcomes and the known outcomes in Experiments 1 and 2. It is particularly noteworthy that there were few differences between judgments of prospective responsibility and retrospective responsibility. Judgments of prospective responsibility may be associated with the decision-maker's role more than his or her choice (Hamilton & Hagiwara, 1992; Vincent, 2011). For example, a teacher may be considered responsible for the welfare of students, regardless of whether the teacher chooses to take them outside on a class trip or not. However, past known outcomes may prompt judgments of retrospective responsibility in which people consider the impact of a choice on the outcome. For example, if an accident occurs, a teacher who chose to take a class outside on a trip might be judged to be more responsible for the accident than one who kept them inside.

4.3 Heuristics and deliberation

A final possibility is that moral judgments may be susceptible to framing effects in part because assessments of moral responsibility and blame depend on a mix of deliberative cognitive processes and automatic, heuristic, or emotional processes (Alicke, 2000; Malle, Monroe & Guglielmo 2014). Such dual process accounts of thinking have been advanced in explanations of moral judgments (e.g., Greene, Sommerville, Nystrom, Darley & Cohen 2001; Moore, Lee, Clark & Conway 2011; Gubbins & Byrne 2014; but see Baron & Gürçay, 2017, for counter-evidence). Similarly, the framing effect has been interpreted as arising from a heuristic or short-cut process to make decisions, an automatic process affected by emotional cues associated with loss (Kahneman, 2011; Loewenstein, Weber, Hsee & Welch, 2001; Glockner & Herbold, 2011). More deliberative, controlled processes may over-ride the framing effect (De Martino, Kumaran, Seymour and Dolan, 2006; Kahneman & Frederick 2007; Gilovich, Griffin & Kahneman, 2002; Whitney, Rinehart & Hinson, 2008; but see Igou & Bless, 2007; Kuo, Hsu & Day, 2009). Hence, the idea that moral judgments are made, at least some of the time, by relying on heuristic or emotional processes may explain why they exhibit a framing effect. In contrast, non-moral judgments about cause and control and counterfactual alternatives may not be as susceptible to framing effects. They may tend to evoke deliberative processes (McEleney & Byrne 2006; Sloman & Lagnado ,2005), and if so, the framing effect may be overridden for such judgments.

Nonetheless, judgments of moral responsibility and blame are influenced by a consideration of whether the agent caused the outcome and whether the agent had control over the outcome (Shaver, 1985; Weiner, 1995; Pizarro, Uhlman & Bloom, 2003; Schlenker et al., 1994; Tetlock, 2000). Responsibility and blame are also influenced by whether the agent could have changed the outcome (Nario Redmond & Branscombe, 1996; Alicke, Buckingham, Zell & Davis 2008). For example, when people listen to a lawyer suggesting a counterfactual about an attack in which changes to the victim's behavior change the outcome, they ascribe higher blame to the victim and lower blame to the attacker (Branscombe, Owen, Garstka & Coleman, 1996; N'gbala & Branscombe 1995). Conversely, people do not imagine an alternative to an action that leads to a bad outcome when the action conforms to a moral norm or obligation (McCloy & Byrne 2000; Walsh & Byrne 2007; Alicke, et al 2008; see Byrne 2016 for a review). Decision making in non-moral domains shares many commonalities with decision-making in moral domains (Uttich & Lombrozo, 2010; Cushman & Young, 2011; Ritov & Baron, 1999; Baron & Ritov 2004; Bucciarelli, Khemlani & Johnson-Laird 2008; Zamir 2014). It is also noteworthy that moral and non-moral judgments were correlated with each other in both experiments. Conversely, there was no echoing effect even for emotion judgments about relief or upset. Hence it may be unlikely that the observation of an echoing effect for moral judgments, and none for non-moral judgments, arises because of a difference between them in heuristic and deliberative processes.

4.4 Conclusions

In everyday life, people's decisions are affected by whether the choices they are presented with are risky or sure options. The experiments we report show that their judgments about the morality of other people's decisions are also affected by whether the decision-maker chose a sure option or a risky option. The results show a moral echoing effect that may have widespread implications for moral judgments in every day life. It suggests that a decision-maker will be credited for a good outcome when he or she made the typical choices of the sure-thing in a gain frame or the risk in a loss frame, and he or she will be discredited when there is a bad outcome and he or she made the atypical choices of a risk in a gain frame or a sure-thing in a loss frame.

5 References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*(3), 368–378.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*(4), 556–474.

Alicke, M. D., Buckingham, J., Zell, E., & Davis, T. (2008). Culpable control and counterfactual reasoning in the psychology of blame. *Personality and Social Psychology Bulletin, 34*, 1371–1381.

Ames, D. L., & Fiske, S. T. (2013). Outcome dependency alters the neural substrates of impression formation. *NeuroImage*, *83*, 599–608.

Baron, J. & Gürçay, B. (2017). A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. *Memory & Cognition*. In press.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of personality and social psychology*, *54*(4), 569–579.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*(2), 74–85.

Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning. *Judgment and Decision making*, *3*(2), 121–139.

Branscombe, N. R., S. Owen, T. A. Garstka, & J. Coleman. (1996). Rape and accident counterfactuals: Who might have done otherwise and would it have changed the outcome? *Journal of Applied Social Psychology, 26*(12), 1042–1067.

Burger, J. M. (1981). Motivational biases in the attribution of responsibility for an accident: A meta-analysis of the defensive-attribution hypothesis. *Psychological Bulletin, 90*(3), 496–512.

Byrne, R.M.J. (2016). Counterfactual Thought. *Annual Review of Psychology, 67*, 135–157.

Chaikin, A. L., & Darley, J. M. (1973). Victim or perpetrator?: Defensive attribution of responsibility and the need for order and justice. *Journal of Personality and Social Psychology*, *25*(2), 268–275.

Cushman, F. & Young, L. (2011). Patterns of moral judgment derive from non-moral psychological representations. *Cognitive Science, 35*(6), 1052–1075.

Darley, J. M., & Shultz, T. R. (1990). Moral rules - their content and acquisition. *Annual Review of Psychology, 41*(1), 525–556.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science, 313*(5787), 684–687.

Druckman, J. N., & McDermott, R. (2008). Emotion and the framing of risky choice. *Political Behavior, 30*(3), 297–321.

Fagley, N.S., Coleman, J.G. , Simon, A.F. (2010). Effects of framing, perspective taking, and perspective (affective focus) on choice. *Personality and Individual Differences, 48*(3), 264–269

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human perception and performance*, *1*(3), 288–299.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

Glöckner, A., & Herbold, A. K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98.

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, *106*(3), 1339–1366.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108.

Gubbins, E., & Byrne, R. M. (2014). Dual processes of emotion and reason in judgments about moral dilemmas. *Thinking & Reasoning, 20*(2), 245–268.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834.

Hamilton, V. L., & Hagiwara, S. (1992). Roles, responsibility, and accounts across cultures. *International Journal of Psychology, 27*(2), 157-179.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*(3), 311–327.

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

Johnson-Laird, P. N., Khemlani, S.S. & Goodwin, G.P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*. In press.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J. P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological review*, *106*(1), 62–88.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Frederick, S. (2007). Frames and brains: elicitation and control of response tendencies. *Trends in cognitive sciences*, *11*(2), 45–46.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society, 47*(2), 263–291.

Kuo, F. Y., Hsu, C. W., & Day, R. F. (2009). An exploratory study of cognitive effort involved in decision under Framing-an application of the eye-tracking technology. *Decision Support Systems, 48*(1), 81–91.

Loewenstein, G. F., Weber, E.U., Hsee, C. K. & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*(2), 267–286. Malle, B. F., Monroe, A. E., & Guglielmo, S. (2014). A theory of blame. *Psychological Inquiry, 25*(2), 147–186.

McCloy, R. & Byrne, R.M.J. (2000) Counterfactual thinking about controllable actions. *Memory & Cognition, 28*(6), 1071–1078.

McCloy, R.A., Byrne, R.M.J. & Johnson-Laird, P.N. (2010). Understanding cumulative risk. *Quarterly Journal of Experimental Psychology. 63*(3), 499–515.

McEleney, A. & Byrne, R.M.J. (2006). Spontaneous causal and counterfactual thoughts. *Thinking and Reasoning, 12*(2), 235–255.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in cognitive sciences, 11*(4), 143-152.

Mitchell, T. R., & Kalb, L. S. (1981). Effects of outcome knowledge and outcome valence on supervisors' evaluations. *Journal of Applied Psychology, 66*(5), 604–612.

Moore, A. B., Lee, N. L., Clark, B. A., & Conway, A. R. (2011). In defense of the personal/impersonal distinction in moral psychology research: Cross-cultural validation of the dual process model of moral judgment. *Judgment and Decision Making*, *6*(3), 186–195.

Nario-Redmond, M. R., & Branscombe, N. R. (1996). It could have been better or it might have been worse: Implications for blame assignment in rape cases. *Basic and Applied Social Psychology*, *18*(3), 347–366.

N'gbala, A. and Branscombe, N.R. (1995) Mental simulation and causal attribution: when simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology, 31*(2), 139–162

Parkinson, M., & Byrne, R. M. (2017). Judgments of Moral Responsibility and Wrongness for Intentional and Accidental Harm and Purity Violations. *Quarterly Journal of Experimental Psychology*. *In press.*

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology, 17*(3), 145–171.

Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology, 39*(6), 653–660

. Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in Judgments of Moral Blame and Praise The Role of Perceived Metadesires. *Psychological Science*, *14*(3), 267–272.

Quelhas, A. C. & Byrne, R.M.J. (2003). Reasoning with deontic and counterfactual conditionals. *Thinking and Reasoning, 9*(1), 43–66.

Rasga, C., Quelhas, A. C., & Byrne, R. M. (2016). Children's reasoning about other's intentions: False-belief and counterfactual conditional inferences. *Cognitive Development*, *40*, 46–59.

Reyna, V. F., Chick, C. F., Corbin, J. C., & Hsia, A. N. (2013). Developmental reversals in risky decision making intelligence agents show larger decision biases than college students. *Psychological science, 25*(1), 76–84.

Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational behavior and human decision processes*, *79*(2), 79–94.

Ritov, I., & Zamir, E. (2014). Affirmative action and other group tradeoff policies: Identifiability of those adversely affected. *Organizational Behavior and Human Decision Processes*, *125*(1), 50–60.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426.

Roszkowski, M. J., & Snelbecker, G. E. (1990). Effects of "framing" on measures of risk tolerance: Financial planners are not immune. *Journal of Behavioral Economics*, *19*(3), 237–246.

Schlenker, B. R., Britt, T. W., Pennington, J., Murphy, R., & Doherty, K. (1994). The triangle model of responsibility. *Psychological review, 101*(4), 632–652.

Sezer, O., Zhang, T., Gino, F., & Bazerman, M. H. (2016). Overcoming the outcome bias: Making intentions matter. *Organizational Behavior and Human Decision Processes*, *137*, 13–26.

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer-Verlag.

Shenhav, A. & Greene, J.D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron, 67*(4), 667–677

Spence, A., & Pidgeon, N. (2010). Framing and communicating climate change: The effects of distance and outcome frame manipulations. *Global Environmental Change*, *20*(4), 656–667.

Sunstein, C.R. (2005). Moral Heuristics. *Behavioral and Brain Sciences, 28*(4), 531–573.

Sloman SA, Lagnado DA. (2005). Do we "do"? *Cognitive Science, 29*, 5–39.

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.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297–323.

Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition, 116*(1), 87–100.

Vincent, N. A. (2011). A structured taxonomy of responsibility concepts. In *Moral Responsibility* (pp. 15–35). Springer. Netherlands.

Viscusi, W. K. (2000). Corporate Risk Analysis: A Reckless Act?. *Stanford Law Review*, 547–597.

Walsh, C. R., & Byrne, R. M. (2007). How people think "if only..." about reasons for actions. *Thinking & Reasoning*, *13*(4), 461–483.

Walster, E. (1966). Assignment of responsibility for an accident. *Journal of Personality and Social Psychology, 3*(1), 73–79.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct.* New York: Guilford Press.

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

Zamir, E. (2014). *Law, Psychology, and Morality: The Role of Loss Aversion*. Oxford. Oxford University Press.

Appendix A: Materials

Experiment 1

Unknown Outcomes

Imagine that John is in charge of the preparations for an outbreak of an unusual flu-like disease, which is expected to kill 600 people in Ireland. John must choose between the two alternative programs to combat the disease that have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

Gain Frame:

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

Loss Frame:

If Program A is adopted, 400 people will die
If Program B is adopted, there is a 1/3 probability that no people will die and a 2/3 probability that 600 will die

Judgments

1. Program A is morally acceptable
2. Program B is morally acceptable

Sure choice: Suppose that John recommends Program A.
Risky choice: Suppose that John recommends Program B.

Gain Frame, Good outcome:

3. John would be morally responsible for people being saved
4. John would be relieved about people being saved
5. John would deserve to be praised for people being saved
6. John would cause people to be saved
7. John would be in control of people being saved

Loss Frame, Good outcome:

3. John would be morally responsible for people not dying
4. John would be relieved about people not dying
5. John would deserve to be praised for people not dying
6. John would cause people to not die
7. John would be in control of people not dying

Gain Frame, Bad outcome

3. John would be morally responsible for people not being saved
4. John would be upset about people not being saved
5. John would deserve to be blamed for people not being saved
6. John would cause people to not be saved
7. John would be in control of people not being saved

Loss Frame, Bad outcome

3. John would be morally responsible for people dying
4. John would be upset about people dying
5. John would deserve to be blamed for people dying
6. John would cause people to die
7. John would be in control of people dying

All conditions

8. Imagine you are a member of a committee formed to review the preparations for the outbreak and to predict how things could turn out. What is the most likely way you would complete the thought, "Things could turn out differently if. . . "
(a) . . . if John recommends the other program
(b) . . . if John were a more moral person
(c). . . if John took more risks
(d). . . if additional programs are proposed other than programs A and B

Experiment 2

Known Outcomes

Imagine that John is in charge of the preparations for an outbreak of an unusual flu-like disease, which is expected to kill 600 people in Ireland. John must choose between the two alternative programs to combat the disease that have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

Gain Frame:

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

Loss Frame:

If Program A is adopted, 400 people will die
If Program B is adopted, there is a 1/3 probability that no people will die and a 2/3 probability that 600 will die

Judgments

1. Program A is morally acceptable
2. Program B is morally acceptable

Sure Choice: Suppose that John recommends Program A.
Risky Choice: Suppose that John recommends Program B.
Gain Frame, good outcome: As a result of his decision, a lot of people were saved.

3. John was morally responsible for people being saved
4. John was relieved about people being saved
5. John deserves to be praised for people being saved
6. John caused people to be saved
7. John was in control of people being saved

Loss Frame, good outcome

As a result of his decision, a lot of people did not die.
3. John was morally responsible for people not dying
4. John was relieved about people not dying
5. John deserves to be praised for people not dying
6. John caused people to not die
7. John was in control of people not dying

Gain Frame, Bad outcome

As a result of his decision, a lot of people were not saved.
3. John was morally responsible for people not being saved
4. John was upset about people not being saved
5. John deserves to be blamed for people not being saved
6. John caused people to not be saved
7. John was in control of people not being saved

Loss Frame, Bad outcome

As a result of his decision, a lot of people died.
3. John was morally responsible for people dying
4. John was upset about people dying
5. John deserves to be blamed for people dying
6. John caused people to die
7. John was in control of people dying

All conditions

8. Imagine you are a member of a committee formed to review the preparations for the outbreak and evaluate how things turned out. What is the most likely way you would complete the thought, "Things could have turned out differently if..."
(a) . . . if John had recommended the other program
(b) . . . if John were a more moral person
(c). . . if John had taken more risks
(d). . . if additional programs had been proposed other than programs A and B

Appendix B: Participants' judgments[3]

A simple analysis shown in Table A1 tested whether participants' own personal judgments had any effect on their other judgments. Here the expected effect of framing is defined in terms of the triple interaction term between gain/loss, sure/risky, and outcome (the three classifications that define the 8 groups of participants), with each variable coded as 1/–1, so that their product is positive when a dependent measure (e.g., blame) is expected to be greater in terms of the usual framing effect. The expected effect of personal judgments is defined as the difference between sure and risky ratings — recall that the participant rated both, in the same outcome condition as the judgment they were evaluating — multiplied by sure/risky, so that the result is positive when the participant's ratings are greater for the option given to the group the participant is in. That is, participants who themselves favored the sure outcome would be expected to assign less blame for the choice of that option.

It is apparent that personal judgments have little or no effect on moral judgments ("Personal" rows), even though these judgments do show the expected framing effect ("Framing" rows) in terms of this simple analysis. They do affect the judgments of being upset by the outcome, but this is not subject to the usual framing effect. (And it would still not be significantly subject to framing if the "personal" effect were included as a predictor, where it could be a nuisance variable.)

[3] We are grateful to Jon Baron for carrying out this informative analysis and reporting it here.

Table A1: Correlations of main dependent variables with expected framing effect and expected personal influence.

	Responsibility	Blame	Upset	Cause	Control
Experiment 1 (r>.137 for p<.05 2-tailed)					
Framing	0.238	0.249	−0.048	0.168	0.123
Personal	0.106	0.037	0.172	−0.001	−0.044
Experiment 2 (r>.150 for p<.05 2-tailed)					
Framing	0.144	0.268	−0.073	0.160	0.147
Personal	−0.107	0.011	0.150	−0.080	−0.086

Appendix C: Statistical analysis of non-moral judgments in Experiment 2

Causal judgments. There was a main effect of outcome, F(1, 165)=33.5, p<.001, η_p^2=.17 as the decision-maker was judged to have caused the good outcome more than the bad outcome, an interaction of the three variables, F(1,165)=4.78, p<.03, η_p^2=.028, and of choice and outcome F(1,165)=4.61, p<.05, η_p^2=.027, and no other differences (largest F=2.77, smallest p < .098). The three-way interaction arises because the decision-maker was judged to have caused a good outcome more than a bad outcome when choosing the risk in a loss frame F(1,165)=17.79, p<.001, η_p^2=.097, or in a gain frame F(1,165) = 12.77, p<.001, η_p^2=.07, or the sure thing in a gain F(1,165)=14, p<.001, η_p^2=.08. No other contrasts showed significant differences on the Bonferroni corrected alpha of .004 (largest F=5.78, smallest p =.017). Causal judgments did not show a framing effect, there was no interaction of frame and choice for good outcomes, F(1,83) =1.57, p=.21, and no interaction of frame and choice for bad outcomes, F(1,82) =3.23, p=.08,as Figure 2D shows. Causal judgments correlated with moral responsibility and blame/praise judgments as reported in the text, and with control r=.45, p<.001, relief/upset judgments r=18, p<.02 and counterfactual judgments r=.19, p<.015.

Control judgments There was a main effect of outcome F(1,165) = 5.84, p<.05, η_p^2=.034 as the decision-maker was judged more in control of good outcomes than bad outcomes, an interaction of the three variables, F(1,165)=3.84, p<.052, η_p^2=.023, and no other differences (largest F=3.19, smallest p<.076). The interaction arises because the decision-maker is judged more in control of a good outcome than a bad outcome when choosing the sure-thing in a gain frame F(1,165)=8.03, p<.005, η_p^2=.05. No other contrasts showed significant differences (largest F=4.6, smallest p<.03). Control judgments did not show a framing effect for good outcomes as there was no interaction of frame and choice for good outcomes, F<1, but when participants knew the outcome was bad there was a framing effect, as the interaction of frame and choice for bad outcomes shows, F(1,82)=5.1, p<.027, η_p^2=.059, see Figure 2E. Control judgments correlated with moral responsibility, blame/praise judgments as reported in the text and with cause as reported above, but not with relief/upset judgments r=–.07, p<.34 or counterfactual judgments r=14, p<.06.

Relief/upset judgments There were no effects of outcome, frame or choice, F<1. There were no framing effects for good-outcomes or bad-outcomes, F<1 in both cases, as Figure 2F shows. Relief/upset judgments correlated only with causal judgments as reported above, they did not correlate with moral responsibility, blame/praise judgments or control also as reported above, or counterfactual judgments r=-.001, p<.99.

Counterfactuals There was no main effect of outcome but outcome interacted with choice, F(1,165)=4.72, p<.05, η_p^2=.02, and choice interacted with frame F(1,165)=3.79, p<.053, η_p^2=.02 and there were no other differences (largest F=2.23, smallest p<.14). Contrasts to decompose the outcome by choice interaction showed that participants agreed that things could have turned out differently for a bad outcome more when the decision-maker chose the sure thing than the risk, F(1,165)=6.78, p<.01, η_p^2=.04, and there were no other significant differences (largest F=3.54, smallest p<.062). Contrasts to decompose the frame by choice interaction showed that partcipants agreed that things could have turned out differently more when the decision-maker chose the sure thing rather than the risk in a loss frame F(1, 165)=6.26, p<.016, η_p^2=.04, and there were no other differences (largest F=3.74, smallest p<.055). Counterfactual judgments did not show a framing effect for bad outcomes, there was no interaction of frame and choice for bad outcomes, F<1, as Figure 2G shows, but there was a framing effect for good outcomes, as the interaction of frame and choice for good outcomes shows, F(1,83)=5.23, p=.03, η_p^2=.06. Counterfactual judgments correlated with moral responsibility and causal judgments as reported above, but not with blame/praise, control, or relief/upset judgments also as reported above.